



Etude théorique des mouvements internes de grande amplitude de la décaalanine et du fragment C-terminal de la protéine ribosomale L7/L12

Yves-Henri Sanejouand

► To cite this version:

Yves-Henri Sanejouand. Etude théorique des mouvements internes de grande amplitude de la décaalanine et du fragment C-terminal de la protéine ribosomale L7/L12. Biophysique [physics.bio-ph]. Université Paris Sud - Paris XI, 1990. Français. NNT: . tel-00266537

HAL Id: tel-00266537

<https://theses.hal.science/tel-00266537>

Submitted on 24 Mar 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ORSAY

n° d'ordre :

UNIVERSITE DE PARIS-SUD
CENTRE D'ORSAY

THESE

présentée
pour obtenir

Le TITRE de DOCTEUR EN SCIENCE

PAR

Yves-Henri Sanejouand

SUJET: Etude théorique des mouvements internes de grande amplitude
de la décaalanine et du fragment C-terminal de la protéine
ribosomale L7/L12

soutenue le 20 Juin 1990 devant la commission d'examen :

MM. J.YON-KAHN Président
J.DURUP Rapporteur
J.SMITH Rapporteur
R.FOURME
A.PACAULT
O.TAPIA

Je tiens à remercier tout particulièrement Mme Jeannine YON-KAHN qui, il y a quelques années déjà, a décidé de me faire confiance et de me laisser partir là où je voulais aller, c'est-à-dire au pays des "théoriciens", et qui, par la suite, m'a laissé parcourir ce pays en toute liberté, même lorsque je lorgnais du côté de la molécule d'eau ou de la molécule "HONO". Je lui en suis très profondément reconnaissant.

Je remercie M. David PERAHIA, qui a dirigé la partie de ce travail consacrée à la décaalanine et à la méthode de dynamique moléculaire confinée. C'est lui, notamment, qui a eu l'idée du principe de cette méthode et je dois dire que ce fut un plaisir de travailler à partir d'une telle idée. Qu'il trouve ici l'expression de ma gratitude.

Je remercie M. François COLONNA-CESARI, qui a guidé, avec M. David PERAHIA, l'essentiel de mes premiers pas dans le monde de l'informatique made in IBM (son éditeur, son système, son JCL...) et qui a su me transmettre rapidement le plus nécessaire de ses compétences. Je le remercie aussi pour les nombreux témoignages de son amitié, pour les nombreuses et stimulantes discussions que nous avons eues à la Maison des Sciences de l'Homme, pour toutes les idées qu'il m'a fait découvrir, pour tous les gens qu'il m'a fait rencontrer et pour m'avoir, l'hiver de l'année dernière, entraîné à sa suite dans le froid pays de Suède...

Je remercie M. Orlando TAPIA qui, depuis ce fameux hiver, a dirigé la partie de mon travail consacré au "CTF" et avec qui j'ai rédigé la partie du chapitre II concernant la théorie des modes normaux. Je tiens à rendre ici hommage à son intuition et à son sens "physique". C'est à lui, notamment, que je dois d'avoir découvert la méthode de Williams.

Je remercie MM. Jean DURUP et Martin KARPLUS pour l'attention qu'ils ont porté à mon travail depuis son tout début, pour leurs conseils, leurs encouragements, leurs remarques et leurs questions, tout à la fois éclairantes et stimulantes.

Je remercie M. Jeremy SMITH pour avoir accepté de faire un rapport sur mon travail, pour les questions qu'il m'a posées, les remarques qu'il m'a faites, et l'intérêt qu'il a manifesté à cette occasion pour mon travail.

INTRODUCTION

Je remercie MM. Adolphe PACAULT et Roger FOURME pour l'honneur qu'il me font en acceptant de juger ce travail.

Enfin, je tiens à remercier l'ensemble des membres du laboratoire d'Enzymologie Physico-Chimique et Moléculaire, ainsi que les habitués du sous-sol de la Maison des Sciences de l'Homme, pour leur amitié, leur bonne humeur, et le plaisir qu'il y a eu à travailler en leur compagnie. Je tiens à remercier tout particulièrement pour cela Mlle Lilianne MOUAWAD et M. Maurice JOUSSET qui ont été, trop souvent peut être (?), confinés dans le même sous-espace que moi.

CHAPITRE I. CINETIQUE DE MOUVEMENTS DE GRAD DE AMPLITUDE DE LA DISPERMANINE

A. Mouvement

B. Calcul de la probabilité de la détection

1. Construction de la distribution et calcul des moments de son énergie potentielle

2. Etude de la qualité de la normalisation effectuée

3. Propriétés propres de la distribution

C. Détection des configurations de répartition

D. Trajectoires de dynamique rétrograde

E. Comparaison des méthodes de dynamique moléculaire moderne et traditionnelle

TABLE DES MATIERES

INTRODUCTION

6

CHAPITRE I: MODELE et METHODES

16

A. Les équations du mouvement

17

B. Modèle physique

19

C. Théorie des modes normaux

22

1. Présentation générale

22

2. Méthode GF de Wilson

27

3. Projection de Williams

29

4. Options pratiques choisies

32

5. Options pratiques et résultats des précédentes études

34

6. Cohérence de l'approche empirique

35

D. Dynamique moléculaire

37

E. Dynamique moléculaire confinée

41

F. Conditions initiales des simulations

43

G. Rotation globale

46

CHAPITRE II: ETUDE DES MOUVEMENTS DE GRANDE AMPLITUDE

DE LA DECAALANINE

48

A. Modèle étudié

48

B. Calcul des axes propres de la décaalanine

50

1. Construction de la decaalanine et minimisation

51

de son énergie potentielle

50

2. Etude de la qualité de la minimisation effectuée

52

3. Fréquences propres de la décaalanine

57

C. Détection des changements de conformation

58

D. Trajectoires de dynamiques confinées

61

E. Comparaison des méthodes de dynamique moléculaire confinée et traditionnelle

64

CHAPITRE III: ETUDE DES MOUVEMENTS DE GRANDE AMPLITUDE D'UNE PETITE PROTEINE, LE FRAGMENT C-TERMINAL DE LA PROTEINE RIBOSOMALE L7/L12

A. Comparaison des surfaces d'énergie potentielle calculées	76
par les programmes CHARMM et GROMOS	79
1. Topologie	79
2. Charges partielles	79
3. Milieu environnant	80
4. Calcul de l'énergie électrostatique	80
B. Minimisation	81
C. Calcul des fréquences et des axes propres	82
D. Première simulation	83
1. Conditions générales de l'étude	83
2. Conditions initiales	84
3. Analyse de la trajectoire obtenue	84
a. Comparaison de la structure moyenne et de la structure cristallographique	86
i) Positions moyennes	86
ii) Fluctuations de la position des atomes	87
b. Comportement individuel des hélices	88
i) Définition dynamique des hélices	88
ii) Structure moyenne des hélices	89
c. Comportement collectif des hélices	90
i) Hélices αA et αC	90
ii) Hélices αA et αB	90
iii) Hélices αB et αC	91
E. Simulation incluant les effets électrostatiques à longue distance	92
1. Analyse de la trajectoire	93
a. Comparaison de la structure moyenne et de la structure cristallographique	95
i) Positions moyennes	95

ii) Fluctuations des positions atomiques	96
b. Propriétés moyennes des hélices	96
c. Mouvements relatifs des hélices	96
d. Conclusion	97
2. Calcul du mouvement des hélices par la théorie des modes normaux	99
a. Calcul classique	99
b. Calcul à partir d'états "excités"	101
CONCLUSION	107
REFERENCES	112

Les protéines, et plus particulièrement les enzymes, jouent un rôle majeur dans les processus vitaux. Les enzymes sont des catalyseurs spécifiques qui permettent aux êtres vivants de fabriquer les molécules nécessaires à leur croissance et à leur reproduction à partir des molécules et des sources d'énergie, solaire, chimique, ou thermique, présentes dans leur environnement. Par exemple, un être unicellulaire comme la levure de bière peut, grâce à une douzaine d'enzymes, transformer du glucose en éthanol avec un rendement très élevé et stocker, sous forme chimique, une partie de l'énergie contenue dans chaque molécule de glucose. En absence d'enzymes, la transformation spontanée d'une molécule de glucose en une molécule d'éthanol est extrêmement lente.

Pour transformer le glucose en éthanol, douze enzymes de la levure de bière catalysent douze réactions chimiques bien définies, chaque enzyme accélérant une réaction chimique qui modifie le produit d'une réaction chimique catalysée par un autre enzyme. Ces douze enzymes forment une véritable chaîne qui, parmi les millions de réactions chimiques possibles, trace un chemin menant du glucose à l'éthanol. Grâce à la spécificité de chacun des enzymes de la chaîne, la transformation du glucose en éthanol s'opère avec un très haut rendement.

En solution, les enzymes ont une structure tridimensionnelle bien définie. Depuis 1960, date à laquelle Kendrew et Perutz publièrent les structures de la myoglobine (J.C.Kendrew & col., 1960) et de l'hémoglobine (M.F.Perutz & col., 1960), obtenues à partir de l'analyse des taches de diffraction de rayons X par des cristaux, plusieurs centaines de structures de protéines ont été obtenues par cette méthode, dont certaines avec une résolution d'environ 1 Å. Les premières études détaillées de structures d'enzymes liés à des molécules analogues à leur substrat naturel permirent de proposer des mécanismes expliquant comment certains d'entre eux fonctionnent. Dans les deux décennies qui suivirent ces succès initiaux, la structure d'une protéine, telle qu'elle est obtenue par la méthode d'analyse de la diffraction des rayons X par des cristaux, fut de plus en plus considérée comme l'information de base autour de laquelle doivent s'articuler les interprétations des autres données expérimentales.

Cette information centrale est une sorte de photographie de la position moyenne de chacun des atomes d'une protéine. L'analyse de la qualité de cette image tridimensionnelle permet d'estimer l'amplitude des fluctuations de la position des atomes autour de leur position moyenne. Cette information sur l'ensemble des positions possibles des atomes dans une structure à une température donnée est quantifiée par les facteurs "B" cristallographiques. A température ordinaire, ces facteurs B décrivent notamment l'amplitude des mouvements des atomes de chacune des molécules du cristal. Cependant, ils n'en décrivent que les mouvements les plus fréquents. En effet, le temps de pose nécessaire pour obtenir l'image tridimensionnelle d'une protéine est très long, plusieurs heures, comparé aux durées caractéristiques des mouvements les plus rapides des atomes d'une protéine, qui sont de l'ordre de grandeur de la femtoseconde (10^{-15} sec) et les changements de conformation rares, et de brève durée, contribuent trop peu à la densité électronique moyenne des atomes d'une structure pour modifier de manière significative la valeur des facteurs B.

Ces changements de conformation rares peuvent être amples. Dès 1954, Hvidt et Linderstrom-Lang ont montré que les cinétiques d'échange de nombreux protons des liaisons peptidiques des protéines sont beaucoup plus lentes que celles des protons de petits peptides (A.Hvidt & K.U.Linderstrom-Lang, 1954). Pour qu'un proton soit échangeable, il faut que la liaison hydrogène dans laquelle il est éventuellement impliqué se brise et qu'il soit alors accessible au solvant (S.W.Englander, 1975). Or, certains protons s'échangent 10^{10} fois plus lentement que d'autres (S.W.Englander & J.J.Englander, 1978). Cela veut dire que les conformations pour lesquelles l'échange est possible sont rares, mais aussi que la forme de la protéine fluctue beaucoup (K.U.Linderstrom-Lang & J.A.Schellman, 1959). Des liaisons hydrogènes se font et se défont et, de temps en temps, les cavités dans l'arrangement compact des atomes à l'intérieur d'une protéine se réarrangent, formant des canaux par lesquels de petites molécules peuvent transiter. Par exemple, ainsi que l'ont montré des expériences d'atténuation de fluorescence (J.R.Lackowicz & G.Weber, 1973), contrairement à l'acrylamide, l'oxygène peut atteindre chacun des tryptophanes de 14 protéines, y compris les plus enfouis d'entre eux.

Grâce à ces réarrangements transitoires, l'oxygène peut rejoindre et quitter ses sites de fixation, enfouis à l'intérieur de protéines telles la myoglobine. L'inaccessibilité au solvant de la poche hémique de la myoglobine est tout à fait nécessaire au bon fonctionnement de cette protéine. En effet, dans la myoglobine, l'atome de fer est sous forme ferreuse Fe^{2+} . Si cet atome entraînait en contact à la fois avec une molécule d'oxygène et avec une molécule d'eau, il s'oxyderait, sous forme Fe^{3+} , et la myoglobine deviendrait incapable de fixer l'oxygène. En fait, l'analyse de la structure de la myoglobine montre que l'arrangement des atomes autour du fer est si compact que même une molécule d'oxygène ne peut atteindre son site de fixation sans rencontrer sur son chemin des barrières énergétiques de plus de 100 Kcal/Mole (M.F.Perutz & F.S.Matthews, 1966; D.A.Case & M.Karplus, 1979). Si la myoglobine était un solide rigide, l'oxygène mettrait plusieurs milliards d'années pour entrer ou sortir de la protéine. Une analyse théorique détaillée montre que les mouvements des chaînes latérales de trois acides aminés, une histidine, une thréonine, et une valine de l'hélice E, suffisent pour abaisser ces barrières énergétiques et permettre le passage de l'oxygène. Enfin, l'analyse de simulations de dynamique moléculaire montre que l'oxygène, passant d'une cavité interne à une autre cavité interne, diffuse hors de la myoglobine en moins d'une nanoseconde (M.Karplus & J.A.McCammon, 1986).

Cette plasticité des protéines joue un rôle majeur dans l'expression de leur fonction. L'importance de cette propriété des protéines est particulièrement évidente dans le cas d'enzymes dont la structure est organisée en domaines, chacun des domaines contenant des résidus participant à la fonction catalytique. Par exemple, dans le cas de la citrate synthase, le plus petit des domaines de la protéine pivote d'environ 18° autour d'un axe proche du résidu 274 lors de la fixation du substrat, isolant ainsi le site actif du solvant (S.J.Remington & col., 1982; C.Chothia & A.M.Lesk, 1984). Grâce à ce mouvement charnière, le site actif de la citrate synthase est, tout d'abord, accessible aux réactifs, puis, durant la catalyse, protégé du solvant. Ce mouvement permet donc de faire pénétrer à l'intérieur de la protéine des molécules bien plus grosses que l'oxygène. De tels déplacements amples et semi-rigides de régions entières d'une protéine lors de l'acte catalytique ont été mis en évidence dans de nombreux autres cas, tels celui de l'hexokinase (C.M.Anderson & col., 1978; W.S.Bennett & T.A.Steitz, 1980), de l'alcool deshydrogénase (H.Eklund & col., 1981), etc.. Dans le cas de

la phosphoglycérate-kinase, étudiée intensivement dans notre laboratoire, le rapprochement de ses deux domaines permet le contact des deux substrats de la réaction que cet enzyme catalyse. Ces modifications de la conformation des enzymes lors de l'acte catalytique peuvent être moins spectaculaires. Par exemple, très souvent, la protéine s'adapte au substrat qui lui est proposé (D.E.Koshland, 1958; 1960). Dans le cas du lysozyme, les changements conformationnels sont localisés au voisinage du site actif (D.C.Philipps, 1966).

Parfois, certaines régions d'une protéine sont si flexibles qu'à l'instar des molécules d'eau non liées elle n'apparaissent pas sur les cartes de densité électronique obtenues par l'analyse des taches de diffraction des rayons X. C'est le cas, par exemple, d'un domaine entier des immunoglobulines, la partie Fc. Cette extrême flexibilité du domaine Fc des immunoglobulines permet à ces protéines de s'adapter à la disposition variable des sites antigéniques à la surface des cellules (R.Huber & W.S.Bennett, 1983). C'est aussi le cas d'une région du trypsinogène, appelée le "domaine d'activation". La flexibilité de cette région doit jouer un rôle important lors de l'activation du trypsinogène. En effet, lorsque l'inhibiteur pancréatique de la trypsine se fixe sur le trypsinogène, ce dernier devient capable de fixer des dipeptides analogues au dipeptide N-terminal de la trypsine, Ile-Val, et le domaine d'activation se rigidifie, devenant visible sur les cartes de densité électronique (R.Huber & W.S.Bennett, 1983).

Le domaine d'activation du trypsinogène reste invisible lorsqu'on analyse la diffraction des rayons X par des cristaux maintenus à très basse température. La grande flexibilité apparente de cette région est donc due au grand nombre de conformations différentes qu'elle peut adopter. Des expériences récentes semblent indiquer que cette forme de flexibilité est une propriété commune à l'ensemble des protéines. Par exemple, la valeur des facteurs cristallographiques B de la myoglobine, extrapolés pour une structure à température nulle à partir des facteurs B obtenus pour des structures étudiées à des températures comprises entre 80 et 300°K, n'est que d'environ deux fois inférieure à la valeur des facteurs B obtenus pour une structure étudiée à une température de 300°K (H.Frauenfelder & col., 1979). Ce résultat signifie qu'une partie des fluctuations de la position des atomes de la

myoglobine à basse température n'est pas due à leurs mouvements à l'intérieur de chacune des molécules du cristal mais plutôt aux diverses positions possibles pour chacun des atomes de la myoglobine dans les différentes conformations présentes dans le cristal. Par ailleurs, certaines de ces expériences suggèrent qu'aux basses températures (de quelques degrés Kelvin à environ 180°K) une protéine ne peut plus subir certains changements de conformation (H.Frauenfelder & col., 1988). Ces expériences mettent en évidence des propriétés analogues à celles d'un mélange de plusieurs protéines. Par exemple, pour de basses températures, la cinétique de reformation du complexe myoglobine-CO n'est pas une fonction exponentielle du temps (R.H.Austin & col., 1975). Lorsque la température baisse, une protéine adopte donc des conformations diverses, dont les énergies libres doivent être très voisines puisque ces conformations sont présentes dans les cristaux dans des proportions équivalentes, mais qui sont isolées les unes des autres par des barrières énergétiques. Selon la terminologie de H.Frauenfelder, chacune des classes de structures ainsi définies est un sous état-conformationnel. La notion de sous-état conformationnel n'intervient ni dans la description des propriétés d'un solide, ni dans celle d'un liquide. Par contre, elle est utilisée pour décrire certaines propriétés des verres (M.Goldstein, 1969). Des propriétés communes aux protéines et à ces liquides solidifiés que sont les verres ont récemment été mises en évidence comme, par exemple, la métastabilité à basse température (E.T.Iben & col., 1989).

A 300°K, une protéine passe rapidement d'un de ces sous-états conformationnels aux autres. L'analyse des fluctuations de la position de l'atome de fer de la myoglobine, mesurées par spectroscopie Mossbauer, montre que ces transitions se produisent en moins de 10^{-7} sec (la durée de la demi-vie de l'atome de Fe^{57}). Ce qui n'est pas vrai à 180°K. A des températures supérieures à 180°K, le spectre Mossbauer de la myoglobine s'enrichit d'un nouveau pic, dont la largeur et l'intensité augmentent rapidement avec la température. L'analyse de ces données montre que des transitions de durées caractéristiques comprises entre 1 et 100 nsec sont responsables de l'apparition de ce pic supplémentaire (E.R.Bauminger & col., 1983). Corrélativement, les fluctuations de la position de l'atome de fer de la myoglobine, qui se produisent en moins de 10^{-7} sec, sont 2,5 fois plus importantes à 300°K (0,25 Å) qu'à 180°K (0,1 Å). L'ordre de grandeur de

la durée caractéristique d'une transition entre deux sous-états conformationnels doit donc être de 10 nsec (F.Parak & W.Knapp, 1984). Comparée à la durée caractéristique des changements de conformation fonctionnellement importants, de l'ordre de grandeur de la seconde dans certains cas, la durée de ces transitions entre sous-états conformationnels peut paraître brève. Cependant, comparée à la période des mouvements de vibration les plus rapides des atomes d'une protéine, environ 0,01 psec (10^{-14} sec), la durée caractéristique de ces transitions entre sous-états conformationnels est très importante. Pour illustrer ces comparaisons, déplaçons ces durées vers des échelles de temps qui nous sont familières. Admettons, par exemple, qu'un mouvement de vibration rapide, tel celui des atomes d'une liaison covalente, dure le temps d'un battement de cœur, soit environ une seconde. En ce cas, une transition entre deux sous-états conformationnels se produit toutes les 10^6 sec, c'est-à-dire tous les 12 jours, et une protéine se replie en 10^{14} sec, c'est-à-dire en 3 millions d'années.

Le rôle fonctionnel de ces transitions entre sous-états conformationnels reste à préciser. A ce jour, on sait seulement que lorsque ces transitions n'ont pas lieu les mouvements des atomes sont de très faible amplitude. Ces transitions doivent donc jouer un rôle important lors de l'expression des fonctions liées à la flexibilité et la plasticité des protéines. Cependant, aujourd'hui encore, on ne peut préciser la nature d'un sous-état conformationnel, ni ce qui différencie deux sous-états conformationnels distincts. A fortiori, on ne peut pas non plus décrire comment une protéine passe d'un sous-état à un autre, ou ce que sont les composantes des barrières énergétiques qu'elle ne peut franchir à des températures inférieures à 180°K.

L'étude de ces transitions entre sous-états peut se faire tant expérimentalement que théoriquement. Les deux approches ont leurs avantages et leurs inconvénients. L'approche théorique a un avantage majeur : elle permet d'étudier les propriétés d'une molécule isolée dans un milieu et dans des conditions parfaitement contrôlés. Cette approche est donc par nature idéale pour aborder l'étude d'un sous-état conformationnel donné et l'étude des transitions possibles à partir de ce sous-état. Cependant, ce type d'approche repose sur de

nombreuses approximations tant théoriques que méthodologiques et la fiabilité de certains résultats est difficile à estimer. En général, un résultat obtenu par une approche théorique n'aura qu'une valeur qualitative et devra être confirmé par des approches indépendantes tant théoriques qu'expérimentales. De leur côté, les approches expérimentales ont un inconvénient majeur : elles ne permettent d'étudier qu'un grand nombre de molécules à la fois. Lorsque des événements complexes ont lieu dans chacune de ces molécules, l'interprétation des résultats obtenus devient délicate. Ces problèmes d'interprétation se rencontrent notamment dans l'étude des phénomènes dont les étapes intermédiaires sont nombreuses et difficiles à caractériser individuellement, comme par exemple, le repliement d'une protéine.

Dans le cas de l'étude des transitions entre les sous-états conformationnels d'une protéine, les approches théoriques devraient pouvoir rapidement apporter une aide à l'interprétation des résultats expérimentaux. En effet, l'ordre de grandeur de la durée caractéristique de ces transitions, 1 à 100 nsec, n'est pas trop éloigné de la durée des plus longues trajectoires de protéine calculées à ce jour via les méthodes de la dynamique moléculaire, 100 psec à quelques nanosecondes. Des trajectoires de quelques nanosecondes sont cependant exceptionnelles, car coûteuses, et, de toute façon, trop courtes. Idéalement, pour confronter les résultats théoriques et expérimentaux, il faudrait disposer de trajectoires de 100 nsec. Nous nous sommes donc proposés, lors de ce travail de thèse, de développer une méthodologie permettant de calculer des trajectoires de plus longue durée.

L'idée de base guidant ces développements est la suivante : au cours d'une simulation de dynamique moléculaire, les mouvements internes les plus lents d'une protéine sont aussi les plus amples, c'est-à-dire ceux qui déplacent le plus les atomes d'une protéine (S.Swaminathan & col., 1982). D'autre part, les mouvements les plus rapides, comme, par exemple, les vibrations des atomes liés de manière covalente, n'influent probablement pas significativement sur l'amplitude et la fréquence des mouvements amples et lents. Nous avons donc cherché à développer une méthode qui ne prenne en compte que les mouvements internes les plus amples et les plus lents d'une protéine. Au cours d'une simulation

de dynamique moléculaire, la lenteur des mouvements internes amples d'une protéine est toute relative puisque leur durée caractéristique est de l'ordre de quelques picosecondes. Cependant, comme la durée des mouvements internes les plus rapides est d'environ 0,01 psec, une méthode qui ne prendrait en compte que les mouvements internes amples et lents d'une protéine au cours d'une simulation devrait être jusqu'à cent fois plus efficace que les méthodes usuelles qui prennent en compte sans distinction aucune tous les mouvements internes possibles. Une telle méthode permettrait donc d'étudier directement les transitions entre les sous-états conformationnels d'une protéine, et de caractériser chacun de ces sous-états. En perspective, l'étude des structures des différents sous-états d'une protéine pourrait servir de base au développement d'une autre méthode ne simulant que les événements les plus importants d'une trajectoire, c'est-à-dire les passages d'un sous-état à un autre. Cette méthode de seconde génération permettrait d'explorer rapidement l'ensemble des possibilités conformationnelles d'une protéine dans son état natif et serait un outil idéal d'aide à l'interprétation des données expérimentales.

Dans la première partie de notre travail, nous présentons le problème qu'aborde la dynamique moléculaire dans toute sa généralité, c'est-à-dire l'étude des mouvements des atomes d'une protéine représentée par un modèle physique. Les propriétés structurales et dynamiques des protéines sont modélisées par une fonction énergie potentielle d'interaction aussi générale et rigoureuse que possible. Les mouvements internes des protéines peuvent alors être complètement décrits et compris dans un cas très particulier : lorsque les déplacements des atomes sont si petits que le comportement dynamique de la protéine est comparable à celui d'un solide. Le mouvement de chacun des atomes peut alors paraître erratique, pourtant, il résulte de la superposition d'un grand nombre de vibrations indépendantes, dont les fréquences sont les fréquences de résonance de la protéine considérée comme un solide. C'est pourquoi on parle d'approximation harmonique, ou d'approximation des modes normaux, en lieu et place d'approximation des petits déplacements. On pourrait aussi parler d'approximation solide. La théorie qui en découle est présentée dans la partie I de notre manuscrit; un peu dispersée dans la littérature, nous espérons, O.Tapia et moi-même, l'avoir présentée ici sous sa forme la plus complète et la plus cohérente. Certains aspects de cette théorie sont mal connus et sont

appliqués dans ce travail pour la première fois au cas des protéines. La comparaison entre cette théorie et les résultats des expériences numériques de dynamique moléculaire est faite dans la partie II par le biais d'une méthode originale, mise au point par D.Perahia et moi-même (Y.H.Sanejouand & col., 1990). Cette méthode, dite de dynamique moléculaire confinée, est le premier stade du développement d'une méthode plus complexe permettant de calculer des trajectoires de longue durée. Cette méthode repose en partie sur la description que la théorie des modes normaux donne des mouvements internes amples des molécules. Comme cette description ne s'avère pas totalement adéquate, la partie III est consacrée à l'étude d'un mouvement interne ample d'une protéine à température ordinaire (Y.H.Sanejouand & O.Tapia, 1990). Il s'agit du mouvement semi-rigide périodique d'une hélice alpha dans un fragment protéique stable, le fragment C-terminal de la protéine ribosomale L7/L12. Les caractéristiques de ce mouvement ample sont probablement comparables à celles du mouvement semi-rigide des domaines d'une protéine. En effet, il s'agit aussi d'un mouvement de domaines, même si l'un des deux domaines considérés est le plus simple qui soit : une hélice α . Comparé au mouvement de domaines de protéines de grande taille, tel celui de la Phosphoglycérate-Kinase, ce mouvement est rapide -sa période est d'environ 8 psec- et, surtout, aisé à mettre en évidence. L'étude théorique détaillée de ce mouvement périodique de grande amplitude est donc techniquement possible, à partir de l'analyse de trajectoires relativement courtes. Les résultats de cette étude, confrontés aux conclusions des travaux menés avec la méthode de dynamique confinée, permettent de définir des grands axes des développements méthodologiques auxquels nous espérons apporter ultérieurement notre contribution.

A. Les équations du mouvement

Pour étudier les propriétés dynamiques d'une molécule de N atomes à partir d'un modèle physique, il faut résoudre les équations du mouvement de cet ensemble d'atomes. Si cet ensemble a n degrés de liberté, il y a n équations à résoudre, dont la forme la plus générale est celle donnée par Lagrange :

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{r}_i} \right) - \frac{\partial L}{\partial r_i} = 0 \quad i = 1, 2, \dots, n \quad (1)$$

où $L = K - V$. L , K et V sont respectivement : le Lagrangien, l'énergie cinétique et l'énergie potentielle de la molécule. Les coordonnées, $\{r_i\}$, choisies pour décrire la position des atomes dans l'espace, peuvent être quelconques : il peut s'agir d'angles, de distances, etc... Seules deux conditions sont imposées au système qu'elles forment : il doit être complet, c'est-à-dire qu'il doit permettre de positionner chacun des atomes de la molécule, et il doit être non redondant; en d'autres termes, les coordonnées de ce système doivent être indépendantes : une modification de l'une d'elles ne doit affecter aucune des autres. Le choix du système de coordonnées est important. Il conditionne aussi bien le degré de difficulté des équations à résoudre que le degré de difficulté de l'interprétation physique des solutions de ces équations.

Le système regroupant les $3N$ coordonnées cartésiennes est celui dans lequel les équations du mouvement ont la forme la plus simple. En effet, on a, par définition :

$$K = \sum_{i=1}^{3N} \frac{1}{2} m_i \dot{x}_i^2 \quad (2)$$

où m_i est la masse de l'atome dont une des trois coordonnées cartésiennes est la coordonnée x_i . Lorsque V ne dépend pas des vitesses des atomes de la molécule, les équations de Lagrange s'écrivent, en fonction des coordonnées cartésiennes :

$$m_i \ddot{x}_i = - \frac{\partial V}{\partial x_i} \quad i = 1, 2, \dots, 3N \quad (3)$$

ou encore, et ce sont les équations de Newton :

$$f_i = m_i \ddot{x}_i \quad i = 1, 2, \dots, 3N \quad (4)$$

puisque l'on a :

$$f_i = -\frac{\partial V}{\partial x_i}$$

où f_i est l'une des trois composantes de la force s'exerçant sur l'atome dont x_i est l'une des trois coordonnées cartésiennes. On peut encore simplifier les $3N$ équations (4) en introduisant un autre système de coordonnées, $\{\eta_i\}$, telles que :

$$\eta_i = \sqrt{m_i} x_i$$

En fonction de ces coordonnées cartésiennes pondérées, l'énergie cinétique, K , et les équations de Newton du mouvement s'écrivent respectivement, à partir des équations (2) et (4) :

$$K = \sum_{i=1}^{3N} \frac{1}{2} \dot{\eta}_i^2 \quad (5)$$

et :

$$f_i = \ddot{\eta}_i \quad i = 1, 2, \dots, 3N \quad (6)$$

f_i étant cette fois :

$$f_i = -\frac{\partial V}{\partial \eta_i}$$

On verra par la suite l'importance de cette généralisation particulière des équations de Newton. Le sens physique de ces équations, ou des équations (4), est le suivant : dans un repère cartésien, pondéré ou non, de l'espace des configurations à $3N$ dimensions, les propriétés dynamiques d'une molécule sont entièrement déterminées par les caractéristiques géométriques locales de la fonction énergie potentielle, V , lorsque celle ci n'est pas une fonction des vitesses des atomes de la molécule.

Le sens physique des équations de Lagrange, (1), dont les équations de Newton ne sont qu'un cas particulier, est plus difficile à préciser. Ce sont des équations plus complexes et donc, le plus souvent, plus difficiles à résoudre. Elles sont utilisées lorsque le nombre de degrés de liberté, n , de la molécule étudiée est bien inférieur à son nombre de degrés de liberté théorique maximal, $3N$, c'est-à-dire lorsque les déplacements des atomes de la molécule sont soumis à de nombreuses contraintes. Dans ce cas, il y a n équations complexes à résoudre au lieu de $3N$ équations plus simples. Les équations de Lagrange

sont aussi utilisées lorsque les propriétés dynamiques d'un ensemble d'atomes sont décrites plus simplement en fonction de coordonnées non cartésiennes. Ce point sera illustré ci-après.

B. Modèle physique

Idéalement, la fonction énergie potentielle, V , devrait être calculée exactement, c'est-à-dire à partir de l'étude de la fonction d'onde de la molécule. Or, une telle étude est tout à fait hors de portée des méthodes actuelles qui sont limitées à l'étude de molécules ayant au plus quelques dizaines d'atomes. Comme les plus petites des protéines ont près de 500 atomes, V est calculé à partir d'une approximation empirique, dont la forme la plus courante est celle adoptée dans les programmes CHARMM (B.Brooks & col., 1983) et X-PLOR (A.Brunger & col., 1989) que nous avons utilisés dans ce travail. Il s'agit de la forme suivante :

$$V = V_b + V_\theta + V_\phi + V_\omega + V_{Vdw} + V_{elec} \quad (7)$$

où V_b est l'énergie des liaisons covalentes :

$$V_b = \frac{1}{2} \sum k_b (l - l_0)^2$$

V_θ est l'énergie des angles de valence :

$$V_\theta = \frac{1}{2} \sum k_\theta (\theta - \theta_0)^2$$

V_ϕ est l'énergie des angles dièdres :

$$V_\phi = \frac{1}{2} \sum |k_\phi| - k_\phi \cos(n\phi) \quad \text{avec } n = 1, 2, 3, 4, 6$$

V_ω est l'énergie des pseudo-angles dièdres :

$$V_\omega = \frac{1}{2} \sum k_\omega (\omega - \omega_0)^2$$

Ce premier ensemble de termes décrit et synthétise de nombreuses données expérimentales concernant la topologie des molécules contenant les atomes participant à la structure des

macromolécules biologiques : la longueur moyenne des liaisons covalentes (l_0), la valeur la plus fréquente des angles de valence (θ_0), des angles dièdres et pseudo-dièdres (ω_0). En général, ces données sont issues de l'étude de la diffraction par les rayons X de cristaux d'acides aminés ou de peptides. Les constantes de force, k_b , k_θ , k_ϕ et k_ω sont des facteurs pondérant les différents termes. Elles proviennent de l'étude des spectres infra rouges ou Raman de petites molécules, de calculs *ab initio*, etc...

La forme de ces termes est volontairement simplifiée. On sait, par exemple, qu'une forme plus correcte de l'énergie potentielle des liaisons covalentes est une fonction de Morse. Cependant, la forme quadratique décrit bien le fond du puits du potentiel de Morse pour des énergies correspondant aux températures couramment étudiées et comme la fonction de Morse contient un terme exponentiel, qui ralentit le calcul de l'énergie potentielle totale, on lui préfère la forme quadratique.

Ces termes "topologiques" de V sont des fonctions simples de coordonnées un peu particulières : en effet, la valeur de ces coordonnées, longueurs de liaisons ou angles définis par 3 ou 4 atomes, ne change pas lors d'une rotation ou d'une translation rigide de la molécule; c'est pourquoi on dit que ces coordonnées sont de type "interne".

La fonction V est complétée par un ensemble de termes "physiques". V_{elec} est l'énergie d'interaction électrostatique :

$$V_{elec} = \sum \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r d_{ij}}$$

V_{Vdw} est l'énergie d'interaction de Van der Waals :

$$V_{Vdw} = \sum \left(\frac{A_{ij}}{d_{ij}^{12}} - \frac{B_{ij}}{d_{ij}^6} \right)$$

La forme du terme répulsif est d'origine empirique. Dans ce cas aussi, un terme exponentiel décrirait mieux la partie répulsive de ce potentiel.

Notons qu'il n'y a, dans la fonction énergie potentielle décrite ci-dessus, aucun terme ad-

ditionnel pour les atomes participant à des liaisons hydrogènes. En effet, il est généralement admis que les liaisons hydrogènes sont de nature principalement électrostatique (L. Pauling, 1960). Elles peuvent donc être représentées d'une manière adéquate par un choix judicieux des charges électrostatiques partielles, q_i et q_j , et des paramètres de Van der Waals, A_{ij} et B_{ij} , pour l'atome donneur, l'atome d'hydrogène et les atomes du groupe accepteur. Par ailleurs, afin de réduire le coût en temps de calcul d'ordinateur (CPU) de l'estimation de V , les effets dus à la polarisation du nuage électronique des atomes sont négligés (ce sont des effets à "N corps"). Enfin, les effets spécifiquement quantiques des interactions atomiques sont aussi implicitement négligés dans les termes physiques ci-dessus, ne serait-ce que parce que les atomes y sont représentés par des masses ponctuelles portant des charges partielles ponctuelles. On ne peut donc, à partir de termes de cette forme, qu'étudier des phénomènes pour lesquels l'approximation classique est parfaitement justifiée.

Ces termes physiques sont fonctions des distances, d_{ij} , séparant les atomes, c'est-à-dire des coordonnées cartésiennes. La fonction énergie potentielle semi-empirique, V , est donc définie en fonction de deux systèmes de coordonnées de nature différente, un système de coordonnées cartésiennes et un système de coordonnées internes.

En général, les termes composant le gradient de la fonction énergie potentielle, V , c'est-à-dire les forces, sont calculés dans un repère cartésien. Notons que les forces dues aux termes topologiques de V sont des forces à 2, 3 et 4 corps. En effet, lorsqu'un angle de valence, par exemple, est écarté de sa position d'équilibre, la force de rappel qui en résulte s'exerce sur les trois atomes formant l'angle considéré : il s'agit donc d'une force à 3 corps. Cette remarque illustre l'importance du choix d'un système de coordonnées. Les termes topologiques de la fonction énergie potentielle ont été choisis en fonction de leur sens physico-chimique. L'écartement plus ou moins grand d'un angle de valence est une notion intuitive. Pourtant, cette notion simple introduit une complication pratique car les forces qui s'opposent à l'écartement d'un angle de valence doivent obéir aux règles strictes de la mécanique newtonienne, c'est-à-dire ne provoquer ni translation ni rotation globale des trois atomes constituant l'angle de valence considéré. Pour ce faire, lorsqu'un angle

de valence est écarté de sa position d'équilibre, des forces doivent agir sur chacun des 3 atomes en fonction desquels cet angle est défini.

A partir de l'expression du gradient de V dans un repère cartésien, on peut poser les $3N$ équations de Newton qu'il faut résoudre pour étudier les propriétés dynamiques d'une molécule. Ces équations ne peuvent être résolues analytiquement que dans un cadre particulier, celui de la théorie des modes normaux, détaillé ci-dessous.

C. Théorie des modes normaux

1) Présentation générale.

Pour de petits déplacements des $3N$ coordonnées r_i par rapport aux coordonnées r_i^0 de la structure étudiée, l'énergie potentielle V peut être développée en série de Taylor :

$$V = V_0 + \sum_{i=1}^{3N} \left(\frac{\partial V}{\partial r_i} \right) (r_i - r_i^0) + \frac{1}{2} \sum_{i,j=1}^{3N} \left(\frac{\partial^2 V}{\partial r_i \partial r_j} \right) (r_i - r_i^0)(r_j - r_j^0) + \dots \quad (8)$$

Or, comme V est une fonction définie à une constante additive près, on peut poser $V_0 = 0$. De plus, si V_0 est un minimum de V , on a :

$$\frac{\partial V}{\partial r_i} = 0 \quad i = 1, 2, \dots, 3N$$

Enfin, si les déplacements des coordonnées sont suffisamment petits, on peut négliger les termes cubiques, quadratiques, etc., dans (8) et approximer V par :

$$V = \frac{1}{2} \sum_{i,j=1}^{3N} k_{ij} (r_i - r_i^0)(r_j - r_j^0) \quad (9)$$

soit, en langage matriciel :

$$V = \frac{1}{2} \mathbf{r}^t \mathbf{F} \mathbf{r} \quad (10)$$

où \mathbf{F} est la matrice des constantes de force de la molécule, k_{ij} , données par :

$$k_{ij} = \frac{\partial^2 V}{\partial r_i \partial r_j}$$

La théorie des modes normaux (H.Goldstein, 1950) repose sur l'hypothèse que l'énergie potentielle V a la forme décrite par l'équation (10). Cette hypothèse est indépendante du système de coordonnées complet et non redondant choisi pour étudier V . Ce système peut, par exemple, être composé de l'ensemble des angles dièdres d'une protéine (T.Noguti & N.Go, 1983; N.Go & col., 1983; M.Levitt & col., 1983). Notons qu'un tel système n'est pas complet. Lors du choix d'un tel système, des contraintes sont implicitement appliquées sur les coordonnées manquantes (les longueurs des liaisons, par exemple). Mais ainsi, le nombre de degrés de liberté de la molécule et la taille de la matrice \mathbf{F} sont considérablement réduits.

Par ailleurs, l'énergie cinétique, K , peut être mise sous la forme matricielle suivante :

$$K = \frac{1}{2} \dot{\mathbf{r}}^t \mathbf{H} \dot{\mathbf{r}} \quad (11)$$

où \mathbf{H} est la matrice des dérivées secondes de l'énergie cinétique par rapport aux vitesses. Dans un repère cartésien standard de l'espace des configurations, $\mathbf{H} = \mathbf{M}$, où \mathbf{M} est la matrice diagonale des masses de dimension $3N \times 3N$. Dans un repère cartésien pondéré, $\mathbf{H} = \mathbf{I}$, où \mathbf{I} est la matrice unité. La forme de l'équation (11) est très générale : cette équation est valable quel que soit le système de coordonnées complet et non redondant utilisé, sauf si des contraintes dépendant explicitement du temps sont imposées à la molécule (H.Goldstein, 1950).

A partir de (10) et (11), on a :

$$L = \frac{1}{2} \dot{\mathbf{r}}^t \mathbf{H} \dot{\mathbf{r}} - \frac{1}{2} \mathbf{r}^t \mathbf{F} \mathbf{r}$$

et les équations de Lagrange, (1), s'écrivent, si \mathbf{H} n'évolue pas en fonction du temps :

$$\mathbf{H} \ddot{\mathbf{r}} + \mathbf{F} \mathbf{r} = 0 \quad (12)$$

puisque l'on a, comme expression du gradient cinétique du lagrangien :

$$\nabla_{\dot{\mathbf{r}}} L = \mathbf{H} \dot{\mathbf{r}}$$

et, pour le gradient spatial :

$$\nabla_{\mathbf{r}} L = -\mathbf{F} \mathbf{r}$$

L'équation (12) est résolue en cherchant une solution de la forme :

$$\mathbf{r} = \mathbf{A} \mathbf{q} \quad (13)$$

les nouvelles coordonnées \mathbf{q} , définies par la relation (13), étant elles mêmes de la forme :

$$q_k = C_k \cos(\omega_k t + \phi_k) \quad (14)$$

Ce type de solution signifie que chacune des coordonnées r_i oscille autour de sa position d'équilibre r_i^0 , son mouvement résultant de la superposition de $3N$ modes de vibration. C_k , et ϕ_k sont respectivement l'amplitude et la phase à l'origine du mode de vibration k . Ces paramètres dépendent des conditions initiales imposées à la molécule. La matrice \mathbf{A} , qui regroupe les directions des $3N$ axes de l'espace des configurations le long desquels tous les atomes vibrent en phase à une même fréquence, et la matrice diagonale Λ , qui regroupe les carrés de la pulsation, w_k , de chacun des modes de vibration de la molécule, dépendent du contenu des matrices \mathbf{H} et \mathbf{F} . En effet, à partir de (13) et de (14), on a :

$$\ddot{\mathbf{r}} = -\mathbf{A} \Lambda \mathbf{q} \quad (15)$$

En remplaçant (15) et (13) dans (12), on obtient :

$$(\mathbf{F} \mathbf{A} - \mathbf{H} \Lambda) \mathbf{q} = 0$$

soit :

$$\mathbf{F} \mathbf{A} = \mathbf{H} \Lambda$$

ou encore :

$$\mathbf{A}^t \mathbf{F} \mathbf{A} = \mathbf{A}^t \mathbf{H} \Lambda \mathbf{A} \quad (16)$$

Pour résoudre (16), c'est-à-dire pour obtenir les matrices \mathbf{A} et Λ à partir des matrices \mathbf{F} et \mathbf{H} , il est nécessaire de poser une condition supplémentaire. On pose :

$$\mathbf{A}^t \mathbf{H} \mathbf{A} = \mathbf{I} \quad (17)$$

et (16) devient :

$$\mathbf{A}^t \mathbf{F} \mathbf{A} = \Lambda \quad (18)$$

Ainsi, \mathbf{A} est la matrice qui diagonalise simultanément \mathbf{F} et \mathbf{H} . On montre que (17) est une condition d'orthogonalité dans un espace dont le tenseur métrique est \mathbf{H} (H. Goldstein, 1950). Ceci apparaît plus clairement si l'on écrit cette équation sous la forme :

$$\mathbf{A} \mathbf{A}^t = \mathbf{H}^{-1} \quad (19)$$

puisque, par définition, une matrice est orthogonale lorsque sa transposée est aussi son inverse, c'est-à-dire lorsque :

$$\mathbf{A} \mathbf{A}^t = \mathbf{I}$$

La matrice \mathbf{A} est donc "une matrice de transformation linéaire faisant passer d'un système d'axes obliques à un système d'axes orthogonaux, comme le montre le fait que le tenseur métrique, \mathbf{H} , transformé par \mathbf{A} est unitaire" (H. Goldstein, 1950). Dans le nouveau repère, le repère "propre", la molécule est décrite par un nouvel ensemble de coordonnées complet et non redondant, les coordonnées "propres", $\{q_k\}$. Dans ce repère, dont le système d'axes de référence est orthogonal, les énergies potentielle et cinétique ont la forme quadratique la plus simple possible. En effet, à partir de (10) et de (13), on a :

$$V = \frac{1}{2} \mathbf{q}^t \mathbf{A}^t \mathbf{F} \mathbf{A} \mathbf{q}$$

soit, en prenant (18) en compte :

$$V = \frac{1}{2} \mathbf{q}^t \Lambda \mathbf{q} \quad (20)$$

De même, à partir de (11) et (13), on a :

$$K = \frac{1}{2} \dot{\mathbf{q}}^t \mathbf{A}^t \mathbf{H} \mathbf{A} \dot{\mathbf{q}}$$

soit, avec (17) :

$$K = \frac{1}{2} \dot{\mathbf{q}}^t \dot{\mathbf{q}} \quad (21)$$

Ce dernier résultat souligne l'origine physique de la condition d'orthogonalité (17), posée à première vue arbitrairement pour résoudre l'équation (16). En posant (17), on choisit parmi tous les systèmes de coordonnées possibles à partir de l'équation (16), un système dans lequel l'énergie cinétique s'exprime sans aucun coefficient, c'est-à-dire un système dans lequel toutes les coordonnées sont équivalentes, d'un point de vue dynamique. On comprend ainsi pourquoi le système des coordonnées cartésiennes pondérées permet de calculer les modes de vibration d'une molécule plus facilement que les autres (J.Tyson & col., 1971; W.D.Gwinn, 1971). En effet, dans ce repère cartésien particulier, l'énergie cinétique s'exprime déjà sans aucun coefficient (équation 5).

Dans le repère propre, la forme de la fonction énergie potentielle, V , a une propriété importante. En effet, puisque Λ est une matrice diagonale, V ne contient pas, dans ce repère, de termes croisés, c'est-à-dire de termes $q_i q_j$. Autrement dit, un mouvement le long de l'un des axes propres de la molécule n'a aucun effet, aucune influence, sur les mouvements se produisant le long des autres axes. On résume cela en disant que les modes de vibration de la molécule ne sont pas couplés. D'autre part, le long de chacun de ces axes, tous les atomes évoluent en phase à une fréquence donnée, ν_k , avec $\nu_k = \frac{\omega_k}{2\pi}$, selon le mouvement d'un oscillateur harmonique, c'est-à-dire celui d'une masse à l'extrémité d'un ressort. Cependant, comme le mouvement de chacun des atomes de la molécule est une superposition de $3N$ vibrations de ce type, les déplacements des atomes peuvent paraître désordonnés, voire erratiques. Ceci illustre une nouvelle fois l'importance du choix d'un système de coordonnées. En effet, tandis que l'évolution des coordonnées cartésiennes des atomes d'une molécule est complexe et apparemment erratique, celle des coordonnées propres est harmonique, c'est-à-dire régulière et simple à décrire. Le système des coordonnées propres est donc particulièrement bien adapté pour décrire les mouvements des atomes d'une molécule lorsque la théorie des modes normaux est applicable, c'est-à-dire lorsque les déplacements des atomes autour de leur position d'équilibre sont petits.

En pratique, on peut toujours calculer les modes de vibration d'une molécule à partir des matrices **F** et **H** calculées pour une configuration atomique donnée. Ces matrices peuvent elles-mêmes être calculées dans un repère quelconque, cartésien ou non-cartésien. On a vu que la fonction énergie potentielle, *V*, est une fonction de coordonnées cartésiennes et de coordonnées internes. Un système de coordonnées internes peut donc être choisi pour calculer les matrices **F** et **H**. C'est le principe de la méthode "GF" de Wilson.

2) Méthode GF de Wilson.

Dans l'application standard de la théorie des modes normaux à l'analyse des spectres infra-rouges, un ensemble particulier de coordonnées internes *s* est utilisé. Ces coordonnées internes sont reliées linéairement aux déplacements infinitésimaux des coordonnées cartésiennes par rapport aux coordonnées \mathbf{r}^0 de la structure étudiée :

$$\mathbf{s} = \mathbf{B}\mathbf{r} \quad (22)$$

les éléments de la matrice **B** étant calculés par les méthodes classiques (E.B.Wilson & col., 1955). **B** est une matrice de dimensions $3N-6 \times 3N$. En effet, il faut $3N-6$ coordonnées internes ($3N-5$ dans le cas particulier d'une molécule linéaire), pour déterminer de manière univoque, c'est-à-dire sans aucune redondance, les positions des *N* atomes d'une molécule, puisque la donnée d'une longueur de liaison, d'un angle de valence et d'un angle dièdre, soient trois coordonnées de type interne, permet de positionner un atome par rapport à trois autres dont les positions sont déjà connues.

Soit **T** une matrice inverse généralisée de **B**, c'est-à-dire une matrice de dimensions $3N \times 3N-6$ définie par :

$$\mathbf{B}\mathbf{T} = \mathbf{I}$$

où **I** est la matrice unitaire de dimensions $3N-6 \times 3N-6$. **B** étant une matrice rectangulaire, son inverse n'est pas défini de manière unique. On a (P.Pulay, 1969) :

$$\mathbf{T} = \mathbf{W}\mathbf{B}^t(\mathbf{B}\mathbf{W}\mathbf{B}^t)^{-1} \quad (24)$$

où B^t est la matrice transposée de B et W une matrice non singulière quelconque de dimensions $3N \times 3N$. Puisque les coordonnées de rotation et de translation ne sont pas explicitement incluses dans s (P.Pulay, 1977), il faut poser :

$$W = M^{-1}$$

où M est la matrice diagonale des masses atomiques, lorsque les coordonnées r sont les coordonnées cartésiennes standards, ou :

$$W = I$$

lorsque les coordonnées r sont les coordonnées cartésiennes pondérées, puisque ces choix permettent de maintenir constantes la position du centre de masse et l'orientation de la molécule (P.Pulay, 1969). On peut alors écrire :

$$r = Ts \quad (23)$$

c'est-à-dire :

$$r = M^{-1}B^t(BM^{-1}B^t)^{-1}s$$

lorsqu'on utilise les coordonnées cartésiennes standards (B.L.Crawford & W.H.Fletcher, 1951). Cette formule est à la base d'une méthode itérative de dynamique moléculaire qui permet de contraindre une ou plusieurs coordonnées internes d'une molécule (D.J.Tobias & C.L.Brooks III, 1988).

A partir de (10), (11) et (23), on peut exprimer l'énergie potentielle, V , et l'énergie cinétique, K , en fonction des coordonnées internes, soit respectivement :

$$V = \frac{1}{2}s^t T^t F T s$$

et :

$$K = \frac{1}{2}\dot{s}^t T^t H T \dot{s}$$

d'où l'on tire l'expression de la matrice des constantes de force dans le repère interne, F_I , de dimensions $3N-6 \times 3N-6$:

$$F_I = T^t F T \quad (24)$$

et où l'on reconnaît la matrice \mathbf{G} de la méthode GF de Wilson (E.B.Wilson, 1939; 1941) :

$$\mathbf{G}^{-1} = \mathbf{T}^t \mathbf{H} \mathbf{T} \quad (25)$$

On peut donc écrire, à l'instar de Wilson :

$$V = \frac{1}{2} \mathbf{s}^t \mathbf{F}_I \mathbf{s} \quad (26)$$

et :

$$K = \frac{1}{2} \dot{\mathbf{s}}^t \mathbf{G}^{-1} \dot{\mathbf{s}} \quad (27)$$

La forme de ces équations est la même que celle des équations (10) et (11). La méthode GF de Wilson est donc bien un cas particulier de la théorie générale des modes normaux présentée ci-dessus. Les axes propres d'une molécule, décrits en termes de fluctuations de coordonnées internes, sont regroupées dans la matrice \mathbf{A}_I qui diagonalise à la fois les matrices \mathbf{F}_I et \mathbf{G}^{-1} .

3) Projection de Williams.

Dans la partie I-B, nous avons vu que l'énergie potentielle V est une fonction de deux types de coordonnées de nature différente, des coordonnées de type cartésien et des coordonnées de type interne. Les équations (22) et (23) montrent que le passage d'un de ces systèmes de coordonnées à l'autre est aisé, au moins en première approximation. Suivant le type d'information que l'on souhaite obtenir, l'un ou l'autre de ces systèmes doit être utilisé. Les coordonnées internes sont très utiles pour décrire les axes propres d'une molécule. En effet, d'une part cette description est indépendante de la position du centre de masse et de l'orientation dans l'espace des axes d'inertie de la molécule, et, d'autre part, il s'agit d'une représentation physique immédiate. Des variations de longueur de liaison, d'angles de valence, d'angles dièdres ou pseudo-dièdres, ont plus de sens que des variations de coordonnées cartésiennes. Par contre, les directions des axes ainsi décrits ne sont indépendantes qu'au voisinage de \mathbf{r}^0 , c'est-à-dire pour des structures très peu différentes de celle pour laquelle la matrice \mathbf{G}^{-1} a été calculée. En effet, ces axes ne sont orthogonaux

que lorsque l'égalité, obtenue en remplaçant \mathbf{H} par \mathbf{G}^{-1} dans (19) :

$$\mathbf{A_I A_I^t} = \mathbf{G} \quad (28)$$

est vérifiée. Or, la matrice \mathbf{G} , tout comme les matrices \mathbf{B} et \mathbf{T} , dépend étroitement de la configuration des atomes de la molécule étudiée. L'équation (28) signifie donc que les coordonnées propres définies à partir des coordonnées internes \mathbf{s} ne forment un système non redondant que très localement; on ne peut donc utiliser ce système particulier de coordonnées à des fins générales, d'analyse ou autres. A l'opposé, les directions des axes propres d'une molécule sont définies dans un repère cartésien de la manière la plus générale possible. En effet, dans un tel repère, la matrice \mathbf{H} ne dépend pas de la géométrie de la molécule. Dans le repère cartésien pondéré on a même : $\mathbf{H} = \mathbf{I}$ et donc :

$$\mathbf{A A^t} = \mathbf{I} \quad (29)$$

Ainsi, la matrice \mathbf{A} est orthogonale au sens usuel du terme et la relation (13) décrit une simple rotation dans un espace de dimension $3N$. D'un point de vue pratique, ce cas particulier est le plus simple à traiter puisque la matrice \mathbf{A} est obtenue en diagonalisant la seule matrice \mathbf{F} .

Notons que le passage par un repère interne est en toute rigueur nécessaire lorsque la matrice \mathbf{F} est calculée dans un repère de type cartésien. Les contributions des champs de force dont l'origine est extérieure à la molécule, c'est-à-dire ceux qui induisent une rotation ou une translation globale de la molécule, sont ainsi éliminées de la matrice \mathbf{F} . Ce passage par le repère interne est d'autant plus nécessaire que de telles contributions sont présentes même lorsque \mathbf{F} est calculée pour une configuration atomique dont l'énergie potentielle est minimale et que la molécule est dans le vide, c'est-à-dire soumise à son seul champ de forces. De telles contributions sont artefactuelles. Elles apparaissent parce que tout déplacement d'un atome, même infinitésimal, provoque une translation et une rotation globale des axes d'inertie de la molécule, c'est-à-dire parce que le repère cartésien est extérieur à la molécule. Or, le calcul d'une constante de force nécessite la détermination de la variation infinitésimale de la force lors de la variation infinitésimale d'une coordonnée. Dans le repère cartésien, ce taux de variation est mésestimé car une partie du déplacement

infinitésimal de l'atome ne modifie pas les distances interatomiques. En absence de toute correction, le bruit ainsi introduit dans le calcul de \mathbf{F} perturbe la détermination des valeurs et des vecteurs propres de cette matrice. Comme conséquence la plus générale, on observe que les six valeurs propres qui devraient être nulles, puisque dans le repère cartésien la molécule est décrite avec six coordonnées de plus qu'il n'en est rigoureusement nécessaire, ne le sont pas tout à fait. Par exemple, Williams (I.H. Williams, 1983) rapporte que dans le cas de H_2O , lors d'un calcul *ab initio* avec une géométrie optimisée, la pseudo-fréquence propre correspondant à la plus grande valeur absolue de ces pseudo valeurs propres nulles est d'environ $i4 \text{ cm}^{-1}$ (la notation "i" indique que la valeur propre correspondante est négative).

Lorsque les plus basses fréquences propres de la molécule sont de l'ordre de grandeur des plus grandes pseudo-fréquences propres nulles, leur détermination est fortement perturbée. De plus, les déplacements des atomes le long des axes propres correspondants induisent alors une rotation et une translation globale de la molécule, faible mais non négligeable.

Pour écarter ces sources d'erreur, et obtenir ensuite une matrice \mathbf{A} orthogonale, il suffit d'exprimer \mathbf{F}_I dans le repère cartésien pondéré. A partir de (22) et (26), on a :

$$V = \frac{1}{2} \mathbf{r}^t \mathbf{B}^t \mathbf{F}_I \mathbf{B} \mathbf{r}$$

et on peut poser :

$$\mathbf{F}_P = \mathbf{B}^t \mathbf{F}_I \mathbf{B} \quad (30)$$

où la matrice \mathbf{F}_P obtenue est bien une matrice de dimensions $3N \times 3N$. Cette matrice est l'expression dans le repère cartésien de la matrice des constantes de force "exactes". L'opération, combinaison des équations (24) et (30), qui relie la matrice des constantes de force calculée dans le repère cartésien à cette même matrice nettoyée de toutes composantes dues à sa détermination dans ce dit repère ou à l'influence de champs de force extérieurs à la molécule :

$$\mathbf{F}_P = (\mathbf{T}\mathbf{B})^t \mathbf{F} (\mathbf{T}\mathbf{B}) \quad (31)$$

est une opération de projection et l'opérateur **TB** est un projecteur. On peut en effet aisément vérifier que l'on a : $(\mathbf{TB})^2 = \mathbf{TB}$, ce qui est la définition d'un projecteur.

Notons que l'opérateur **TB** est le projecteur décrit par I.H.Williams (I.H.Williams, 1983). La démonstration ci-dessus montre comment sa méthode de projection est intimement reliée au calcul classique des modes normaux d'une molécule. Physiquement parlant, le projecteur de Williams permet d'extraire de la matrice des constantes de force, **F**, calculée dans le repère cartésien, toutes les composantes dont l'origine n'est pas interne à la molécule. En toute rigueur, cette méthode devrait être systématiquement employée lorsque les modes de vibration d'une molécule sont calculés à partir de matrices **F** et **H** définies dans le repère cartésien.

A ce jour, cette méthode de projection n'a été appliquée qu'à de toutes petites molécules. Dans la dernière partie de cette thèse, nous l'appliquerons à une protéine de près de 600 atomes. Auparavant, nous en étudierons les effets sur une molécule de taille intermédiaire, la décaalanine.

4) Options pratiques choisies.

Pratiquement, plusieurs voies sont possibles pour calculer les directions des axes propres d'une molécule dans un repère cartésien. Quelle que soit la voie choisie, il est toujours intéressant de disposer des directions des axes propres d'une molécule exprimées dans chacun des deux systèmes de coordonnées, cartésien et interne. Pour ce faire, notons qu'on peut relier directement les matrices **A** et **A_I**. En effet, à partir de (18) et (30), on a :

$$\mathbf{A}^t \mathbf{B}^t \mathbf{F}_I \mathbf{B} \mathbf{A} = \Lambda$$

soit :

$$(\mathbf{B} \mathbf{A})^t \mathbf{F}_I (\mathbf{B} \mathbf{A}) = \Lambda$$

ce qui montre que :

$$\mathbf{A_I} = \mathbf{B}\mathbf{A} \quad (32)$$

De même, on a :

$$\mathbf{A} = \mathbf{T}\mathbf{A_I} \quad (33)$$

la matrice \mathbf{A} étant ici de dimensions $3N \times 3N-6$. Ce sont les six directions des axes propres de rotation et de translation qui manquent. Réciproquement, dans (32), la matrice $\mathbf{A_I}$ est de dimensions $3N-6 \times 3N$ et six des colonnes de cette matrice sont entièrement nulles.

Ayant la possibilité de calculer \mathbf{A} à partir de $\mathbf{A_I}$, on peut donc utiliser la méthode la plus classique, c'est-à-dire diagonaliser simultanément les matrices $\mathbf{F_I}$ et \mathbf{G}^{-1} (méthode GF de Wilson). Des programmes existent dans les bibliothèques mathématiques des superordinateurs qui font cela comme, par exemple, F02AEF de la bibliothèque NAG (J.H.Wilkinson & C.Reinsch, 1971). Notons que les calculs de \mathbf{G}^{-1} et $\mathbf{F_I}$ nécessitent la détermination de \mathbf{B} et deux inversions de matrice (équation 24 et 25). On peut cependant calculer directement ces deux matrices par des méthodes de différences finies (M.Levitt & col., 1983). En ce qui concerne $\mathbf{F_I}$, la méthode la plus fiable et la moins coûteuse repose sur le calcul de :

$$\mathbf{k_i} = \frac{\partial \mathbf{f_I}}{\partial s_i}$$

par l'approximation suivante :

$$\mathbf{k_i} = \frac{1}{2\Delta s} (\mathbf{f_I}(s_1, \dots, s_i + \Delta s, \dots, s_{3N-6}) - \mathbf{f_I}(s_1, \dots, s_i - \Delta s, \dots, s_{3N-6}))$$

où $\mathbf{k_i}$ est le vecteur qui regroupe les $3N-6$ constantes de force k_{ij} de la colonne i de la matrice $\mathbf{F_I}$, où Δs est une petite quantité scalaire et où les forces exprimées dans le repère interne, $\mathbf{f_I}$, sont déduites des forces exprimées dans le repère cartésien, \mathbf{f} , au moins en ce qui concerne les termes physiques, par (P.Pulay, 1969) :

$$\mathbf{f_I} = \mathbf{T}^t \mathbf{f} \quad (34)$$

Ayant $\mathbf{F_I}$, on peut aussi calculer $\mathbf{F_P}$ par l'équation (30); cela ne nécessite que le calcul, d'ailleurs très rapide, de \mathbf{B} . Cependant, comme la détermination complète de $\mathbf{F_I}$ nécessite

$2 \times (3N-6)$ calculs du gradient de V , et que l'inversion d'une matrice est de toute façon nécessaire, la méthode la plus rapide doit être la suivante : calculer \mathbf{F} analytiquement dans le repère cartésien (M.Levitt, 1972; H.Katz & col., 1979), puis projeter cette matrice avec l'opérateur de Williams.

Dans ce travail, nous avons choisi une approche plus coûteuse, mais aussi plus générale que cette dernière, puisque nous avons calculé \mathbf{F} dans le repère cartésien pondéré, par une méthode de différences finies. Pour des ordinateurs scalaires, ou qui utilisent mal les possibilités de la vectorisation, le surcoût consécutif au choix d'une méthode de différences finies pour calculer \mathbf{F} est négligeable lors de l'étude de macromolécules biologiques, puisque le temps de CPU nécessaire à une diagonalisation ou à une inversion de matrice $N \times N$ est à peu près proportionnel à N^2 tandis que celui nécessaire au calcul de la matrice \mathbf{F} est proportionnel à N . Cela n'est pas le cas avec le VP200 dont la bibliothèque mathématique contient des sous-programmes particulièrement bien vectorisés qui effectuent les opérations de diagonalisation et d'inversion et le surcoût constaté lors du calcul des modes de vibration d'une petite protéine de 600 atomes est d'environ 30 % du temps CPU, dans l'hypothèse la plus défavorable. Cela dit, projection comprise, un tel calcul ne prend qu'environ vingt minutes de temps CPU. Par contre, la quantité de mémoire requise pour l'effectuer est importante. Par exemple, il faut 26 Méga-octets pour stocker la seule matrice \mathbf{T} lorsque celle-ci est calculée en double précision. Pour la matrice \mathbf{F} , il en faut environ 13 (elle est symétrique). Actuellement, la taille des macromolécules que l'on peut étudier dans le cadre de la théorie des modes normaux est surtout limitée par la quantité de mémoire vive nécessaire pour stocker les matrices calculées.

5) Options pratiques et résultats des précédentes études.

A notre connaissance, la théorie des modes normaux n'a été appliquée qu'à une seule protéine décrite dans un repère cartésien, l'inhibiteur pancréatique de la trypsine bovine (BPTI) (B.Brooks & M.Karplus, 1983; J.Smith & col., 1989) et elle l'a été sans projec-

tion préalable de la matrice \mathbf{F} . Les autres études ont toutes été faites avec un système de coordonnées internes réduit, dans lequel seuls les angles dièdres étaient retenus, en diagonalisant simultanément les matrices \mathbf{F}_I et \mathbf{G}^{-1} correspondantes. C'est ainsi que le BPTI avait été initialement étudié (T.Noguti & N.Go, 1982; N.Go & col., 1983; M.Levitt & col., 1983). Et c'est ainsi que furent aussi étudiés la crambine, la ribonucléase et le lysozyme (M.Levitt & col., 1985). En ce qui concerne les points qui nous intéressent principalement, les résultats de ces travaux sont les suivants : la fréquence du mode de vibration le plus lent est toujours comprise entre 2 et 5 cm^{-1} et les modes de vibration de fréquence inférieure à 30 cm^{-1} sont responsables de l'essentiel des fluctuations de la position des atomes. Ce sont ces modes de vibration que nous étudierons tout au long de cette thèse. Nous nous intéresserons tout particulièrement aux axes de l'espace des configurations le long desquels se font ces vibrations.

6) Cohérence de l'approche empirique.

Les résultats ci-dessus ont été obtenus en étudiant les propriétés dynamiques d'une molécule décrite par un modèle physique purement classique. On peut donc se demander si ces résultats dépendent ou non des aspects quantiques des interactions atomiques. Selon Peticolas, la réponse est non car les modes de vibration dont la fréquence est supérieure à 200 cm^{-1} sont peu, ou pas, peuplés à température ordinaire (W.L.Peticolas, 1979). Cette réponse mérite d'être détaillée. En effet, l'énergie totale moyenne, $\langle E_t \rangle$, d'un oscillateur plongé dans un bain thermique porté à la température T est, suivant la formule de Planck pour un oscillateur harmonique n'ayant qu'un seul degré de liberté :

$$\langle E_t \rangle = \frac{1}{2} h\nu + \frac{h\nu}{e^{\frac{h\nu}{kT}} - 1} \quad (35)$$

ν étant la fréquence de l'oscillateur, k et h , les constantes de Boltzmann et de Planck, respectivement. Lorsque la fréquence de l'oscillateur est petite, ou la température élevée, c'est-à-dire lorsque $\frac{h\nu}{kT} \ll 1$, on peut appliquer l'approximation suivante :

$$e^{\frac{h\nu}{kT}} = 1 + \frac{h\nu}{kT} + \dots$$

et la formule (35) devient la formule bien connue de Boltzmann :

$$\langle E_t \rangle = kT$$

Lorsque $\frac{h\nu}{kT} = 1$, c'est-à-dire, pour une température de 300°K, lorsque la fréquence de l'oscillateur est d'environ 200 cm⁻¹, $\langle E_t \rangle = 1,08kT$, ce qui correspond quantitativement à ce que donnerait la formule classique de Boltzmann, à moins de 10% près. Cependant, le premier terme de la formule (35), $\frac{1}{2}h\nu$, qui est celui de l'énergie de point 0, contribue pour 0,5 kT à ce résultat. Le comportement d'un oscillateur de fréquence 200 cm⁻¹ dans un bain thermique à 300°K est donc largement gouverné par le caractère quantique des interactions atomiques. A cette même température, lorsque la fréquence de l'oscillateur est de 20 cm⁻¹, le terme d'énergie de point 0 ne contribue plus que pour 5% à l'énergie totale moyenne de l'oscillateur, qui est alors égale à kT, à moins de 0,1% près. L'étude classique de modes de vibration de fréquence inférieure à 20 cm⁻¹ est donc raisonnable et les mouvements de grande amplitude se développant dans les protéines doivent pouvoir être correctement décrits et étudiés par des outils physiques classiques.

Une étude d'un polypeptide en hélice α , la décaglycine, a même montré que dans le cadre des approximations de la théorie des modes normaux, la description classique des fluctuations des positions atomiques prédit 94 % des fluctuations exactes, effets quantiques compris, pour des formules théoriques appliquées à une température de 50°K, et 99,8 % de celles-ci à 300°K (R.M.Levy & col., 1982). Cependant, ces calculs furent faits en négligeant notamment les effets physiques du terme d'énergie de point 0 et des couplages entre les mouvements rapides, dont l'évolution est dominée par des effets quantiques, et les mouvements lents, dont l'évolution est correctement décrite par des méthodes classiques. Ce sujet ne peut donc être aujourd'hui considéré comme définitivement clos.

Une autre question majeure est la suivante : les solutions des équations du mouvement des atomes données par la théorie des modes normaux sont-elles une première approximation suffisante des solutions exactes ? Pour y répondre, il faut comparer les solutions que donne cette théorie aux solutions exactes. Ces dernières peuvent être obtenues

numériquement, en résolvant les équations du mouvement (4) pas à pas. Cette méthode, dite de dynamique moléculaire, est détaillée ci dessous.

D. Dynamique moléculaire

Le principe de la résolution itérative des équations du mouvement de Newton (équations 4 ou 6) est le suivant : connaissant les forces, les vitesses, et la configuration des atomes à un instant t , les positions et les vitesses des atomes sont calculées un bref instant plus tard. Ainsi, de proche en proche, on peut décrire la trajectoire des atomes dans le repère cartésien. Un des algorithmes les plus utilisés pour ce faire, le plus simple de surcroît, est celui de Verlet (L.Verlet, 1967) :

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \ddot{\mathbf{r}}\Delta t^2 \quad (36)$$

où Δt est le pas temporel d'intégration (cet algorithme peut être appliqué aussi bien aux coordonnées cartésiennes standards qu'aux coordonnées pondérées). Les vitesses atomiques y interviennent implicitement. En effet, on a en première approximation :

$$\dot{\mathbf{r}} = \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t - \Delta t)}{2\Delta t}$$

Les accélérations des atomes, $\ddot{\mathbf{r}}$, sont obtenues à partir des forces agissant sur les atomes. Or, le calcul des forces est long puisqu'il faut calculer $3N \times (N + 1)/2$ termes, c'est-à-dire l'ensemble des interactions entre les N atomes pris deux à deux, pour obtenir les N forces agissant sur les N atomes. Ce mode de résolution des équations du mouvement est donc limité par la durée du calcul des forces d'interaction. Il est aussi limité par la taille du pas temporel utilisé pour l'intégration des équations. Ce pas est choisi de façon à ce que l'énergie totale du système étudié soit constante lorsque celui-ci est isolé. Pour une protéine, les valeurs usuelles de ce pas sont voisines de 0,001 psec. C'est pourquoi, aujourd'hui encore, malgré l'évolution rapide de la puissance des superordinateurs, on ne calcule pas de trajectoires de protéines de plus de quelques nanosecondes. Encore s'agit-il des plus longues; la plupart des trajectoires de protéines ne dépassent pas 100 psec. Or, 100 psec ne représentent que quelques périodes, 6 à 15, de mouvements vibratoires

de fréquences comprises entre 2 et 5 cm^{-1} , tels les modes de vibration de plus basse fréquence d'une protéine. De plus, les phénomènes biologiques les plus intéressants, tels les remaniements conformationnels accompagnant la fixation d'un substrat ou le repliement d'une protéine, se produisent sur des échelles de temps allant de la microseconde à quelques centaines de secondes (J.A.McCammon & S.C.Harvey, 1987).

C'est pourquoi, afin d'aborder l'étude de ces phénomènes, l'on a cherché à développer des méthodes permettant d'intégrer les équations du mouvement des atomes avec un pas temporel plus élevé. Les valeurs usuelles du pas d'intégration sont fixées dans la gamme de la femtoseconde par la valeur de la fréquence des mouvements les plus rapides qu'il faut intégrer. Dans le cas d'une protéine dans le vide, il s'agit des vibrations des liaisons covalentes O-H et N-H. La période de ces oscillations est d'environ 0,01 psec. On peut donc, en gelant les mouvements vibratoires de haute fréquence, utiliser un pas temporel plus grand pour intégrer les équations du mouvement des atomes d'une molécule. Un tel procédé est même à recommander puisqu'il mime un effet quantique connu. Pour ce faire, il suffit, par exemple, de supposer que les modes de vibration de haute fréquence sont strictement localisés sur les liaisons covalentes et de geler celles-ci tout au long de la trajectoire, ce qui revient à imposer aux équations du mouvement des atomes des contraintes de la forme :

$$\sigma(r_i, r_j, t) = (r_i(t) - r_j(t))^2 - d_{ij}^2 = 0 \quad (37)$$

Les équations du mouvement sont alors résolues en tenant compte de ces conditions qui stipulent que la distance entre les atomes liés i et j doit rester fixe, de longueur d_{ij} , tout au long de la simulation. Dans l'algorithme SHAKE (J.P.Ryckaert & col., 1977), les forces de contraintes sont déterminées successivement, une par une, au cours d'un processus itératif jusqu'à ce que toutes les conditions (37) soient satisfaites. L'utilisation d'un grand nombre de contraintes sur les distances interatomiques appliquées au moyen de cet algorithme est devenu un procédé courant pour réduire le coût en CPU des dynamiques moléculaires. Cela permet d'intégrer les équations du mouvement avec un pas environ deux fois plus grand que le pas usuel. Les effets de ces contraintes sur les propriétés moyennes et dynamiques des protéines sont faibles (W.F.Van Gunsteren & M.Karplus, 1982). Par contre, s'il est aussi possible de contraindre les angles de valence à conserver leur valeur initiale tout au long

d'une trajectoire, l'utilisation combinée de ces deux types de contraintes réduit d'un facteur deux les fluctuations des positions des atomes. De plus, pour des ensembles atomiques compliqués tels des protéines, la méthode de résolution des équations de contrainte (37) ne converge pas toujours. Les méthodes basées sur des algorithmes du type SHAKE ne permettent donc de gagner au mieux qu'un facteur deux sur la durée du calcul d'une simulation.

Mais ceci n'est pas étonnant si l'on tient compte de ce qu'enseigne la théorie des modes normaux lorsqu'elle est appliquée au cas des protéines. La méthode ci-dessus ne peut être efficace que si l'on peut considérer qu'un mode de vibration est localisé sur chacune des liaisons covalentes et sur chacun des angles de valence contraint. Or, ce n'est strictement exact pour aucun des modes de vibration d'une protéine. Pour chacun d'eux, même les plus localisés, tous les atomes oscillent en phase. Cependant, plus la fréquence d'un mode de vibration est élevée, plus il a tendance à être localisé, c'est-à-dire plus le nombre d'atomes qui vibrent avec une amplitude significative à cette fréquence est réduit. Pour une protéine dans le vide, le cas extrême est celui des modes de vibration O-H. En général, seuls un atome O et un atome H liés covalamment vibrent significativement à une des fréquences comprises entre 3400 et 3500 cm^{-1} . Pour de tels modes, la méthode SHAKE donne de très bons résultats.

A l'opposé, les modes de vibration de très basse fréquence sont délocalisés; des segments entiers de la protéine vibrent en phase pour chacun d'eux (N.Go & col., 1983; M.Levitt & col., 1983; B.Brooks & M.Karplus, 1983). Or, comme nous l'avons vu dans la partie B6, les mouvements composés de vibrations de basse fréquence sont les objets d'étude par excellence de la méthode de dynamique moléculaire. Il serait donc souhaitable de n'étudier que ceux là. Un calcul simple montre qu'on devrait pouvoir utiliser un pas de 0,1 psec pour intégrer un mouvement dont les composantes vibrationnelles les plus rapides ont une fréquence de 33 cm^{-1} , puisqu'un pas de 0,001 psec est nécessaire pour intégrer un mouvement dont les composantes vibrationnelles les plus rapides ont une fréquence de 3300 cm^{-1} . Ce calcul donne l'espoir de réaliser, pour un coût en CPU donné, des trajectoires

de protéine 100 fois plus longues que d'ordinaire. Pour ce faire, une méthode se présente à l'esprit : n'intégrer les équations de Newton que le long des axes des modes de vibration de basse fréquence d'une molécule. Si la méthode SHAKE repose sur l'hypothèse que les modes de vibration de haute fréquence sont localisés, ce qui en limite l'efficacité, la méthode cherchée repose sur celle que les directions des axes propres d'une molécule conservent un sens physique pour des structures assez différentes de la structure pour laquelle elles ont été déterminées. Une telle méthode permettrait aussi de confronter directement les solutions exactes des équations du mouvement à celles obtenues à partir de la théorie des modes normaux, et donnerait ainsi une mesure de la validité de ces dernières.

Noguti et Go (T.Noguti & N.Go, 1985) ont développé une méthode de ce type, la méthode SCV (Scaled Collective Variables), mais en l'appliquant au cas de la méthode Monte Carlo. Le principe de la méthode Monte Carlo est le suivant (N.Metropolis & col., 1953) : des configurations atomiques sont choisies au hasard puis leur énergie potentielle est calculée. Suivant la valeur de cette énergie par rapport à l'énergie moyenne attendue à la température étudiée, une configuration est ou non retenue, les énergies des configurations retenues étant réparties suivant une statistique imposée, celle de Maxwell-Boltzmann, en général. Des quantités moyennes peuvent ensuite être calculées pour cet ensemble de configurations, telles les fluctuations des positions atomiques. L'efficacité de cette méthode est limitée par la façon dont sont choisies les configurations atomiques. Pour ce faire, la méthode SCV utilise un pas spatial, collectif et anisotrope. Pour chercher une nouvelle configuration à partir d'une configuration sélectionnée auparavant, les atomes sont déplacés collectivement, c'est-à-dire tous à la fois. Mais, et c'est là l'originalité de cette méthode, les amplitudes des déplacements des atomes ne sont pas choisies entièrement au hasard. A travers une pondération, elles prennent en compte les déplacements qui se produisent corrélativement le long des axes propres de la molécule; ils doivent être petits le long des axes des modes de vibration de fréquence élevée et importants le long des autres. Ceci, parce que le potentiel est moins abrupt, moins raide, le long des axes des modes de vibration de basse fréquence.

Cette méthode a été appliquée au BPTI. Elle s'avère 50 à 500 fois plus efficace qu'une méthode standard. Une telle efficacité démontre que les axes propres définis par la théorie des modes normaux gardent leur signification physique pour des configurations atomiques bien différentes de celle pour laquelle ils ont été déterminés. Notons cependant que les directions des axes propres utilisés par Noguti et Go dans leur travail sont définies en termes d'angles dièdres, c'est-à-dire en fonction de coordonnées de type interne. De plus, elles sont recalculées toutes les 10000 étapes.

Dans le présent travail, nous avons adapté, D.Perahia et moi-même, la méthode SCV au cas de la dynamique moléculaire, qui présente au moins deux avantages sur la méthode Monte Carlo. Tout d'abord, elle offre l'occasion d'étudier des séries temporelles; ensuite, comme l'énergie totale d'une molécule dans le vide doit impérativement se conserver, la méthode de dynamique moléculaire impose un critère rigoureux et sensible à tout développement méthodologique. Le principe de notre méthode, que nous avons appelé méthode de dynamique moléculaire confinée, est le suivant : comme dans le cas de la méthode qui utilise l'algorithme SHAKE, on cherche à geler les mouvements de vibration de haute fréquence; pour ce faire, des contraintes sont imposées le long des axes des modes de vibration de fréquence élevée. Aucun déplacement ne doit s'y produire. Cette méthode se situe donc à mi-chemin entre la méthode reposant sur l'algorithme SHAKE et la méthode SCV de Noguti et Go. Cette méthode est détaillée ci-dessous.

E. Dynamique moléculaire confinée

On peut exprimer les équations de Newton du mouvement des atomes dans le système des coordonnées propres. En effet, on a, à partir de l'équation (13) :

$$\ddot{\mathbf{r}} = \mathbf{A}\ddot{\mathbf{q}} \quad (38)$$

puisque la matrice \mathbf{A} , lorsqu'elle est définie dans le repère cartésien, ne varie pas au cours du temps. Par ailleurs, en développant l'énergie potentielle, V , au premier ordre, respec-

tivement par rapport aux coordonnées propres et cartésiennes, on a :

$$V = V_0 + \mathbf{f}_Q^t \mathbf{q}$$

où \mathbf{f}_Q est le gradient de V dans le système des coordonnées propres, et :

$$V = V_0 + \mathbf{f}^t \mathbf{r}$$

où \mathbf{f} est le gradient de V dans le système des coordonnées cartésiennes. Soit, en incluant (13) dans l'équation ci-dessus :

$$V = V_0 + \mathbf{f}^t \mathbf{A} \mathbf{q}$$

d'où, en comparant les deux développements :

$$\mathbf{f}_Q = \mathbf{A}^t \mathbf{f} \quad (39)$$

Les équations de Newton :

$$\mathbf{f} = \mathbf{H} \ddot{\mathbf{r}}$$

s'écrivent alors, à partir de (38) et (39) :

$$(\mathbf{A}^t)^{-1} \mathbf{f}_Q = \mathbf{H} \mathbf{A} \ddot{\mathbf{q}}$$

en fonction des coordonnées propres. Ou :

$$(\mathbf{A}^t)^{-1} \mathbf{f}_Q = (\mathbf{A} \mathbf{A}^t)^{-1} \mathbf{A} \ddot{\mathbf{q}}$$

puisque : $\mathbf{A} \mathbf{A}^t = \mathbf{H}^{-1}$ (équation 19). Soit, finalement :

$$\mathbf{f}_Q = \ddot{\mathbf{q}}$$

La forme des équations de Newton est donc conservée lors du changement de repère décrit par l'équation (13) lorsque \mathbf{A} est exprimée dans un repère cartésien. Le système des coordonnées propres est dans ce cas un repère d'inertie de l'espace des configurations et l'on peut intégrer les équations de Newton dans ce repère tout aussi bien que dans un repère cartésien. Par ailleurs, on peut résoudre ces équations en présence de contraintes holonomes, c'est-à-dire de contraintes de la forme :

$$\sigma(\mathbf{r}, t) = 0$$

Le principe de la méthode de dynamique moléculaire confinée est d'imposer le long des $3N-P$ axes des modes de vibration dont les fréquences sont les plus élevées des contraintes de la forme :

$$q_i = 0$$

ou encore :

$$\ddot{q}_i = 0$$

En pratique, cela se passe ainsi : une étape de projection est ajoutée à l'algorithme standard de la dynamique moléculaire, juste avant l'algorithme de Verlet. A partir du gradient de V calculé dans le repère cartésien pondéré, \mathbf{f} , on calcule le gradient de V , \mathbf{f}_Q , dans le sous-espace dans lequel la dynamique moléculaire est confinée. Ce sous-espace est défini par les P axes propres sélectionnés, c'est-à-dire ceux le long desquels aucune contrainte n'est imposée. On a, à partir de (39) :

$$\mathbf{f}_Q = \mathbf{A}_P^t \mathbf{f}$$

où \mathbf{A}_P^t est la sous-matrice de \mathbf{A}^t de dimensions $P \times 3N$. Puis, à partir de \mathbf{f}_Q , on calcule le gradient de V dans le repère cartésien pondéré, \mathbf{f}^P :

$$\mathbf{f}^P = \mathbf{A}_P \mathbf{f}_Q$$

où \mathbf{A}_P est la sous-matrice de \mathbf{A} de dimensions $3N \times P$. On vérifie que la combinaison des deux opérations précédentes :

$$\mathbf{f}^P = \mathbf{A}_P \mathbf{A}_P^t \mathbf{f}$$

est une opération de projection puisque l'on a bien :

$$(\mathbf{A}_P \mathbf{A}_P^t)^2 = \mathbf{A}_P \mathbf{A}_P^t$$

F. Conditions initiales des simulations

Les équations du mouvement ne déterminent pas complètement le comportement dynamique d'une molécule. Il faut leur adjoindre des conditions initiales, c'est-à-dire une

structure et un ensemble de vitesses initiales. Dans la plupart des simulations de dynamique moléculaire du présent travail, la structure initiale est celle pour laquelle ont été calculés les axes propres de la molécule, c'est-à-dire une configuration atomique pour laquelle l'énergie potentielle a une valeur minimale. De ce fait, à $t=0$, toute l'énergie de la molécule est sous forme cinétique.

Traditionnellement, une vitesse initiale est attribuée à chaque atome de façon à ce que l'énergie cinétique initiale totale soit nettement inférieure à celle correspondant à la température moyenne de la simulation à effectuer. La molécule est ensuite progressivement chauffée au cours d'une période dite de thermalisation. Cette période peut durer de 10 à 20 picosecondes. Cette façon de déterminer les conditions initiales d'une simulation est lourde et peu reproductible puisque, en pratique, ce genre de détails méthodologiques (mode de répartition au hasard des vitesses, mode et vitesse du chauffage progressif, etc...) est rarement précisé dans la littérature. D'autre part, lorsque l'approximation de la théorie des modes normaux est applicable, l'énergie cinétique présente à $t=0$ sur un axe propre donné est aussi l'énergie totale présente sur cet axe tout au long de la dynamique, puisqu'il n'y a aucun couplage entre les mouvements se produisant le long d'axes propres distincts. Dans un tel cas, répartir au hasard l'énergie cinétique initiale sur les différents axes propres d'une molécule revient à favoriser certains types de mouvements. Or, comme, à énergie totale constante, l'amplitude du mouvement d'un oscillateur harmonique est inversement proportionnelle à sa fréquence, les fluctuations de la position des atomes peuvent donc dans un tel cas varier énormément d'une simulation à l'autre, en fonction des conditions initiales choisies.

Cependant, il faut préciser que lors de dynamiques moléculaires réalisées à partir d'une fonction énergie potentielle empirique, des couplages existent entre les mouvements se produisant le long de différents axes propres et l'énergie cinétique se répartit peu à peu plus uniformément, en moyenne, sur les différents axes propres d'une molécule. Notons cependant que la durée d'un tel processus d'équilibration peut être très importante.

C'est pourquoi nous avons adopté une méthode originale d'initialisation qui consiste à répartir uniformément l'énergie cinétique initiale sur les axes propres d'une molécule. Soit T la température moyenne désirée pour la simulation et P le nombre de degrés de liberté de la molécule. En supposant que la température initiale est double de la température moyenne, on a, pour l'énergie cinétique initiale, K_I :

$$K_I = Pk_B T$$

où k_B est la constante de Boltzmann. C'est-à-dire, en fonction de chaque coordonnée propre, q_i :

$$\frac{1}{2} \dot{q}_i^2 = k_B T \quad i = 1, \dots, P$$

ou encore :

$$\dot{q}_i = \sqrt{2k_B T} \quad i = 1, \dots, P$$

A partir de cet ensemble de vitesses initiales le long des axes propres on peut déterminer les vitesses cartésiennes initiales des atomes, puisque l'on a, d'après l'équation (13) :

$$\dot{\mathbf{r}} = \mathbf{A} \dot{\mathbf{q}} \quad (41)$$

Cette méthode d'initialisation, qui est un cas particulier de la méthode proposée par Sloane et Hase (C.S.Sloane & W.L.Hase, 1977), a au moins deux avantages. Tout d'abord, d'un point de vue théorique, elle est idéale puisqu'elle permet d'obtenir des valeurs moyennes parfaites à partir d'une seule simulation, lorsque le comportement dynamique de la molécule est correctement décrit par la théorie des modes normaux, c'est-à-dire en première approximation. Ensuite, comme on le vérifiera dans la dernière partie de cette thèse, cette méthode d'initialisation permet d'éviter toute période de thermalisation. Ainsi, une macromolécule peut avoir dès la première étape d'une simulation une énergie cinétique aussi élevée qu'on le souhaite.

A ce jour, des méthodes d'initialisation appartenant à cette famille n'ont été employées, à notre connaissance, que pour de très petites molécules, de quelques atomes chacune. Dans la dernière partie de cette thèse, nous l'emploierons pour calculer des trajectoires

d'une molécule de près de 600 atomes, le fragment C-terminal de la protéine ribosomale L7/L12 d'E.Coli.

RESULTATS

G. Rotation globale.

La méthode de dynamique moléculaire confinée présentée dans le chapitre I-E n'est pas adaptée au cas d'une molécule qui serait en rotation constante ou même à celui d'une molécule qui serait soumise à un champ de forces externes qui ferait fluctuer son moment de rotation. En effet, l'orientation dans l'espace des axes propres de la molécule est indépendante du temps, puisque ces axes sont définis dans un repère cartésien. Ils ne conservent donc leur sens physique que si les axes d'inertie de la molécule ne sont pas en rotation par rapport au repère cartésien.

Théoriquement, à la distribution de vitesses initiales (41) doit correspondre un moment de rotation global nul pour la molécule, si la température initiale est nulle sur l'ensemble des axes propres de rotation. Par la suite, la molécule étant isolée, c'est-à-dire non soumise à un champ de forces extérieur, son moment de rotation doit se conserver, c'est-à-dire rester nul, et la molécule ne doit pas tourner. Pratiquement, du fait des erreurs d'intégration, dues notamment à l'imprécision des algorithmes, le moment de rotation global d'une molécule fluctue au cours d'une trajectoire et la molécule peut tourner sur elle-même petit à petit. L'importance de cette rotation peut être estimée en comparant les structures finales et initiales d'une trajectoire. Cette comparaison se fait en ajustant les deux structures par la méthode des moindres carrés (avec une pondération par les masses) après avoir superposé leurs centres de masse. On peut ainsi définir un axe et un angle de rotation global (W.Kabsch, 1976). Cet angle est calculé lors de chacune de nos simulations. En pratique, pour les trajectoires de dynamique moléculaire confinée réalisées dans le présent travail, qui sont toutes assez courtes (10 psec au plus), cet angle reste négligeable (inférieur à un degré). La molécule ne tourne donc jamais sur elle-même de manière significative au cours des trajectoires dont nous allons présenter ci-dessous les analyses.

II. Etude des mouvements de grande amplitude de la décaalanine.

A. Modèle étudié

Le modèle moléculaire que nous avons choisi pour développer et étudier la méthode de dynamique confinée est un polypeptide enroulé en hélice α , la décaalanine, dont les extrémités N- et C-terminales sont chimiquement bloquées, respectivement par un groupe CH_3CO et par un groupe NHCH_3 , ces groupes permettant de neutraliser les charges normalement présentes aux extrémités des peptides en milieu aqueux à pH 7. Les groupes CH et CH_3 sont représentés par des atomes étendus, c'est-à-dire des pseudo-atomes dont les paramètres (masse, charge, etc..) miment ceux de l'ensemble des atomes de ces groupes. Ainsi, la décaalanine a 66 atomes.

L'hélice α est une structure à la fois très répandue dans les protéines, dont les propriétés structurales sont connues avec une grande précision et qui a été beaucoup étudiée tant par des approches expérimentales que théoriques (K.Itoh & T.Schimanouchi, 1970; B.Fanconi & col., 1971; B.Fanconi & W.L.Peticolas, 1971; K.G.Brown & col., 1972). Dans ces deux types d'approches, l'étude des homopolypeptides a été privilégiée. Certaines propriétés de ces molécules sont convenablement décrites par une combinaison d'approches expérimentales et théoriques. Par exemple, une fonction énergie potentielle empirique correctement paramétrisée permet de reproduire les spectres infra-rouge et Raman des homopolypeptides avec une précision de 5 cm^{-1} (S.Krimm, 1983), notamment ceux de la polyalanine (J.F.Rabolt & col., 1977; A.M.Dwivedi & S.Krimm, 1982). Parmi ces homopolypeptides, l' α -décaglycine est un modèle très souvent utilisé lors d'études théoriques de nature fondamentale (R.M.Levy & col., 1982; Y.H.Sanejouand, 1985; D.Perahia & col., 1990) tandis que l' α -décaalanine est un modèle utilisé en de plus rares occasions (T.Simonson, 1990). Pourtant, la glycine est le seul acide aminé qui n'a aucune chiralité. Dans ce travail, nous avons préféré étudier un homopolymère du plus simple des acides aminés possédant une chaîne latérale, et donc une chiralité, l'alanine. La conformation

de cette homopolyalanine, une hélice α , a été choisie en raison de sa stabilité et de sa simplicité. En effet, on peut aisément déterminer les paramètres structuraux d'une hélice α , c'est-à-dire son pas, son rayon et l'angle que font les plans de deux résidus successifs de l'hélice. F.Colonna-Césari a écrit un programme qui calcule les caractéristiques géométriques des structures secondaires d'une protéine tout au long d'une simulation de dynamique moléculaire. Nous avons inclus ce programme dans CHARMM. Pour une hélice α , une des quantités calculées est un angle qui mesure la cambrure de son axe. Le principe du calcul de cet angle est le suivant : l'axe de l'hélice est estimé une première fois à partir de la position des atomes C_α de quatre résidus successifs, i à $i+3$. Il l'est ensuite à partir de la position des atomes C_α des résidus $i+1$ à $i+4$. S'il s'agit d'une hélice α idéale, l'angle que font deux estimations successives de la direction de l'axe de l'hélice est nul. Pour une hélice réelle, la valeur maximum de l'angle que font ces estimations consécutives tout au long de l'hélice est une mesure de la cambrure de l'axe de l'hélice à un instant donné d'une dynamique.

Dans cette partie de notre travail, toutes les interactions électrostatiques et de Van der Waals sont prises en compte lors des calculs d'énergie potentielle et la constante diélectrique, ϵ_r , est posée égale à 1. Ce dernier choix conduit sans doute à surestimer quelque peu les interactions électrostatiques dans la molécule. En effet, la constante diélectrique macroscopique d'un milieu quelconque est due à trois effets microscopiques indépendants dont les deux premiers, la réorientation des dipôles électriques permanents en présence d'un champ électrostatique et les fluctuations de la valeur de ces dipôles, sont implicitement pris en compte lors des simulations de dynamique moléculaire (voir, par exemple, T.Simonson, 1990). Le troisième effet, la polarisation des nuages électroniques des atomes, peut aussi bien réduire qu'augmenter l'intensité de l'interaction électrostatique entre deux atomes proches. Cet effet est très sensible à la configuration des atomes au voisinage des sources d'un champ électrostatique. Pour deux atomes distants de plus de 5-6 Å, une constante diélectrique comprise entre 1,2 et 1,4 rend compte de cet effet de manière convenable (A.Warshel & S.T.Russel, 1984, F.Colonna-Césari & O.Tapia, 1990). Compte tenu de l'imprévisibilité de cet effet aux courtes distances, nous avons préféré le négliger dans la présente étude.

Au total, lors de chaque calcul de l'énergie potentielle de la décaalanine, 1860 termes d'interactions entre atomes non-liés sont évalués. Les 124 interactions entre atomes séparés par trois liaisons covalentes successives (interactions "1-4") sont évaluées avec des charges partielles réduites d'un facteur 2,5 ($E_{14FAC}=0,4$), comme il est d'usage de le faire. Les paramètres utilisés dans la fonction énergie potentielle (charges partielles, constantes de force, etc...) sont ceux de la version 19 de CHARMM.

B. Calcul des axes propres de la décaalanine.

1) Construction de la décaalanine et minimisation de son énergie potentielle.

L'énergie potentielle de la décaalanine construite à partir des valeurs standards des angles Φ ($-57,4^\circ$) et Ψ ($-47,5^\circ$) pour une hélice α est de -328,14 Kcal/Mole. Cette valeur ne correspond pas à un minimum de la fonction énergie potentielle utilisée. A l'issue de la minimisation de cette fonction, au moyen de l'algorithme POWELL (M.J.D.Powell, 1977), l'énergie potentielle de la décaalanine est de -390,19 Kcal/Mole. Le tableau II-1 montre l'évolution des différents termes de l'énergie potentielle au cours de la minimisation. Les termes électrostatiques (29,37 Kcal/Mole) et de Van der Waals (37,83 Kcal/Mole) sont responsables, à part sensiblement égales, de l'essentiel de la différence observée entre les énergies potentielles de la structure idéale et de la structure obtenue à la suite de la minimisation (62,05 Kcal/Mole).

Tableau II-1 : Evolution des termes de l'énergie potentielle de la décaalanine, en Kcal/Mole, au cours de la minimisation de cette fonction. La géométrie de la structure de départ est celle d'une hélice α idéale ($\Phi = -57,4^\circ$, $\Psi = -47,5^\circ$).

V	totale	liaisons	angles	dièdres	impr.	VDW	électr.
ΔV	-62,05	0,41	3,25	1,07	0,40	-37,83	-29,37

Les angles Φ et Ψ de la structure dont l'énergie potentielle a été minimisée ont respectivement pour valeur, en moyenne et pour les résidus de la région centrale de l'hélice, -62° et $-42,5^\circ$. Durant la minimisation, les angles dièdres de la décaalanine varient d'environ 5° . L'amplitude de cette variation est faible mais non négligeable. Cependant, pour chacun des résidus, la valeur de la somme des angles Φ et Ψ reste à peu près constante. En d'autres termes, l'angle que font deux liaisons peptidiques consécutives de la décaalanine varie très peu durant la minimisation (J.A.Tainer & col., 1982). Par contre, l'orientation de la chaîne latérale des résidus change. A la fin de la minimisation, la structure de la décaalanine est un peu plus proche de celle d'une hélice α_I que de celle d'une hélice α_{II} (pour cette dernière, les valeurs idéales des angles Φ et Ψ sont respectivement de $-70,5^\circ$ et $-35,8^\circ$). On montre que la structure obtenue appartient bien à la famille des hélices de type α . En effet, son pas est de $1,50 \pm 0,09 \text{ \AA}$, et l'angle que font deux résidus consécutifs est de $99,4 \pm 2,2^\circ$, pour des valeurs idéales de $1,5 \text{ \AA}$ et 100° , respectivement. Les liaisons hydrogènes entre le groupement C=O d'un résidu i et le groupement N-H du résidu $i+4$ se maintiennent au cours de la minimisation. Il y en a 8 au total. En moyenne, la distance entre l'atome accepteur et l'atome donneur est de $2,9 \text{ \AA}$, ce qui correspond parfaitement à la valeur expérimentale. Par contre, l'angle que font les paires d'atomes (N,H) et (H,O) impliqués dans une liaison hydrogène varie de 17° à 33° , alors qu'une liaison hydrogène dans une hélice α idéale est linéaire. Cette non-linéarité des liaisons hydrogène est la principale différence structurale entre l'hélice que nous étudions et une hélice idéale.

Les différences existant entre cette structure et une hélice idéale peuvent être dues à la simplicité de notre modèle : approximation des atomes étendus pour les chaînes latérales, absence d'effets de polarisation, absence de solvant, constante diélectrique unitaire, etc... De plus, l'hélice que nous étudions est petite; elle n'est enroulée que sur 3 tours (il y a 3,6 résidus par tour dans une hélice α). Or, les paramètres géométriques d'une hélice α idéale sont ceux d'une hélice infinie. Ces différences peuvent aussi être l'indice qu'une erreur affecte certains des paramètres utilisés. En effet, les paramètres de la fonction énergie potentielle de la décaalanine ne sont pas des valeurs *ad hoc*. La structure de la décaalanine en hélice α (et dans le vide) ne fait pas partie des informations qui ont permis de les déterminer. En ce sens, l'étude de ce système permet de jauger la qualité de la partie de

la fonction énergie potentielle de CHARMM qui concerne le résidu alanine avec un atome étendu. Enfin, les différences relevées ci-dessus peuvent être dues au fait que l'hélice que nous étudions est une structure dont l'énergie potentielle a été minimisée, c'est-à-dire un système dont la température est de 0°K, alors que les données expérimentales concernent des systèmes dont la température est de 300°K. Notons que lors de la minimisation de la fonction énergie potentielle de la structure cristallographique d'une petite protéine, la crambine, les angles Φ et Ψ varient pour la plupart respectivement de -5 à -20° et de +5 à +20° (M.Whitlow & M.N.Teeter, 1986). Notre résultat pourrait donc être tout à fait général.

2) Etude de la qualité de la minimisation effectuée.

Il est classique de calculer les axes propres d'une molécule après avoir minimisé sa fonction énergie potentielle (H.Goldstein, 1950). Pratiquement, lors d'une minimisation numérique de nature itérative, l'énergie potentielle de la structure finale obtenue n'est jamais un minimum exact de la fonction énergie potentielle. Nous allons donc, dans un premier temps, étudier l'impact de la qualité de la minimisation effectuée sur la précision de la détermination des axes propres de la décaalanine. Il s'agit de définir des critères d'arrêt du processus itératif de minimisation et de vérifier que la structure de la décaalanine obtenue à l'issue de la minimisation décrite ci-dessus répond à ces critères.

Pour ce faire, une courte trajectoire de la décaalanine à 300°K, de 2.10^{-4} psec, a été calculée avec une méthode de simulation traditionnelle. La durée totale de cette trajectoire est inférieure à la valeur du pas d'intégration le plus usuel d'une simulation de dynamique moléculaire (0,001 psec). A $t=0$, l'énergie potentielle de la décaalanine a une valeur minimale. Les configurations atomiques obtenues à chacune des étapes de cette simulation sont donc comparables à des structures dont la fonction énergie potentielle est de plus en plus mal minimisée. Le principe de notre étude est de chercher des indicateurs sensibles corrélés à la qualité de la minimisation, celle-ci étant supposée diminuer au cours du temps, lors

de notre simulation.

Nous calculons donc, pour des structures prises à différents instants de cette très courte trajectoire, différents indicateurs; tout d'abord, une quantité globale rapide à évaluer et standard, le gradient moyen de l'énergie potentielle de la structure, f :

$$f = \sqrt{\frac{1}{3N} \sum_{i=1}^{3N} \left(\frac{\partial V}{\partial r_i} \right)^2}$$

Durant les $2 \cdot 10^{-4}$ psec de la simulation, f évolue linéairement (Fig. II-1a). A $t=0$, c'est-à-dire pour la structure dont l'énergie potentielle a été minimisée, sa valeur est de $2,2 \cdot 10^{-6}$ Kcal/mole.Å.

Les autres indicateurs suivis lors de cette étude sont plus longs et plus délicats à calculer. Ils permettent de relier les valeurs que prend le gradient moyen à la qualité de la minimisation. Comme la minimisation de la fonction énergie potentielle de la décaalanine est menée à bien dans le but de calculer les fréquences et les axes propres de cette molécule, la qualité de la minimisation est mesurée par la précision avec laquelle sont déterminées ces fréquences et ces axes. On calcule donc les fréquences et les axes propres de la décaalanine en fonction du temps, c'est-à-dire pour des structures qui sont de plus en plus différentes de la structure dont l'énergie potentielle a été minimisée. Pour ce faire, la matrice des constantes de force, \mathbf{F} , est calculée pour chacune d'entre elles. Ensuite, soit cette matrice est directement diagonalisée, et l'on obtient des fréquences et des axes propres dits "bruts", soit elle est au préalable projetée par la méthode de Williams, et l'on obtient des fréquences et des axes propres dits "purs". La projection de Williams est une étape en toute rigueur nécessaire, d'un point de vue mathématique. La présente étude est une occasion de vérifier si cette étape est aussi nécessaire d'un point de vue pratique.

Nous avons vu que la projection de Williams permet de séparer les 6 directions des axes propres de rotation et de translation des autres. En d'autres termes, à la suite de cette projection, les 6 pseudo-fréquences propres associées à ces 6 axes sont exactement nulles.

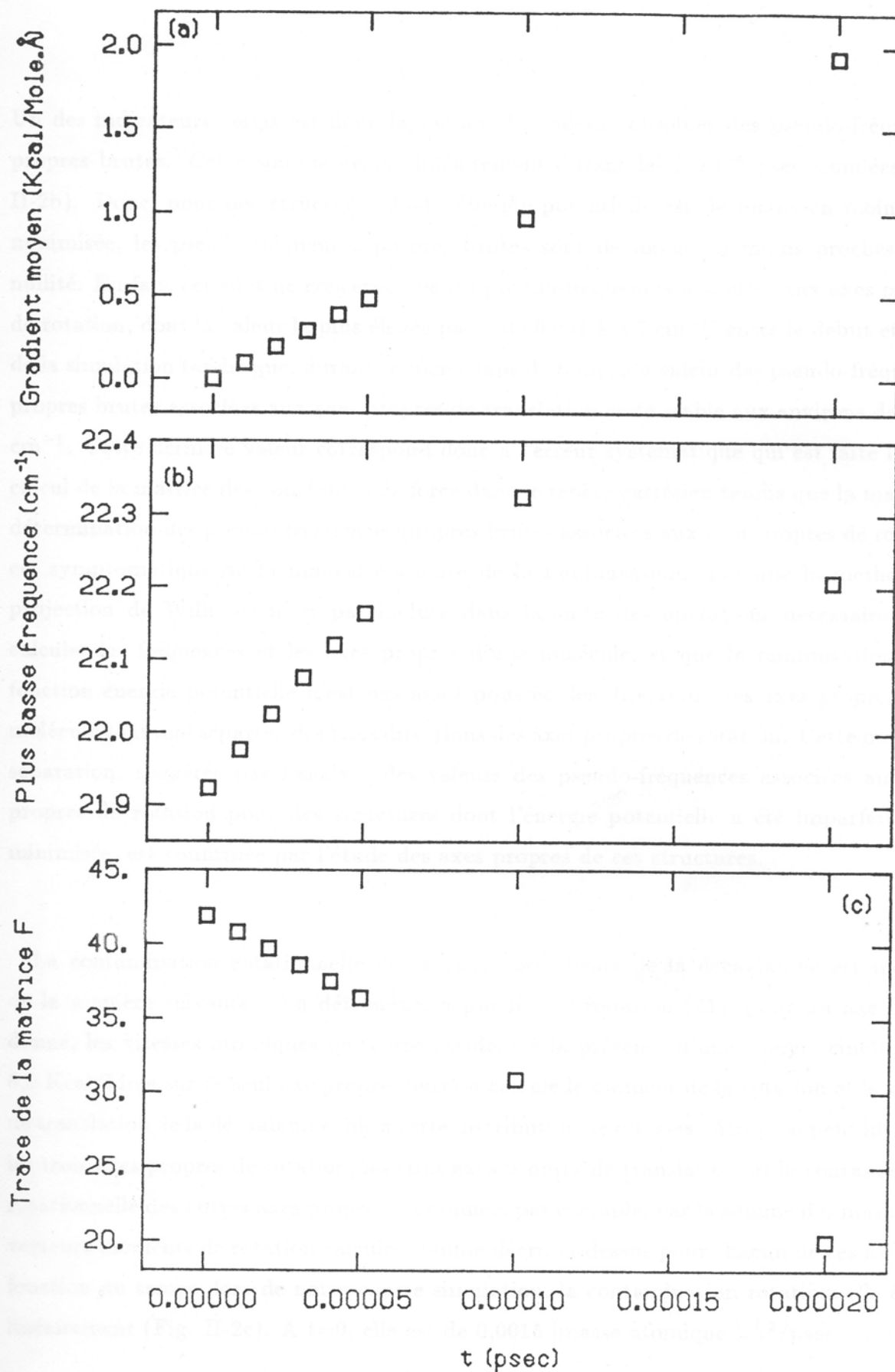


Figure II-1. Etude de structures dont l'énergie potentielle est de plus en plus mal minimisée : (a) gradient moyen; (b) plus basse fréquence propre "pure"; (c) trace de la matrice des constantes de force (en unités arbitraires).

Un des indicateurs suivis est donc la somme des valeurs absolues des pseudo-fréquences propres brutes. Cette somme évolue linéairement durant les $2 \cdot 10^{-4}$ psec simulées (Fig. II-2b). Donc, pour des structures dont l'énergie potentielle est de moins en moins bien minimisée, les pseudo-fréquences propres brutes sont de moins en moins proches de la nullité. En fait, cet effet ne concerne que les pseudo-fréquences associées aux axes propres de rotation, dont la valeur la plus élevée passe de 0,001 à $4,7 \text{ cm}^{-1}$ entre le début et la fin de la simulation tandis que, durant ce même laps de temps, la valeur des pseudo-fréquences propres brutes associées aux axes propres de translation reste stable aux environs de $0,001 \text{ cm}^{-1}$. Cette dernière valeur correspond donc à l'erreur systématique qui est faite lors du calcul de la matrice des constantes de force dans le repère cartésien tandis que la mauvaise détermination des pseudo-fréquences propres brutes associées aux axes propres de rotation est symptomatique de la mauvaise qualité de la minimisation. Lorsque la méthode de projection de Williams n'est pas incluse dans la suite des opérations nécessaires pour calculer les fréquences et les axes propres d'une molécule, et que la minimisation de la fonction énergie potentielle n'est pas assez poussée, les directions des axes propres de la molécule sont mal séparées des trois directions des axes propres de rotation. Cette mauvaise séparation, suggérée par l'analyse des valeurs des pseudo-fréquences associées aux axes propres de rotation pour des structures dont l'énergie potentielle a été imparfaitement minimisée, est confirmée par l'étude des axes propres de ces structures.

La contamination rotationnelle des axes propres bruts de la décaalanine est mesurée de la manière suivante : on détermine, à partir de l'équation (41), pour un axe propre donné, les vitesses atomiques qui correspondent à la présence d'une énergie cinétique de 0,5 Kcal/Mole sur ce seul axe propre, puis on calcule le moment de la rotation et la vitesse de translation de la décaalanine dûs à cette distribution de vitesses. Ainsi, on peut identifier les trois axes propres de rotation, les trois axes propres de translation, et la contamination rotationnelle des autres axes propres est donnée, par exemple, par la somme des normes des vecteurs moments de rotation calculés comme décrit ci-dessus pour chacun de ces axes. En fonction du temps, lors de notre courte simulation, la contamination rotationnelle évolue linéairement (Fig. II-2c). A $t=0$, elle est de $0,0015 [\text{masse atomique}] \cdot \text{\AA}^2/\text{psec}$.

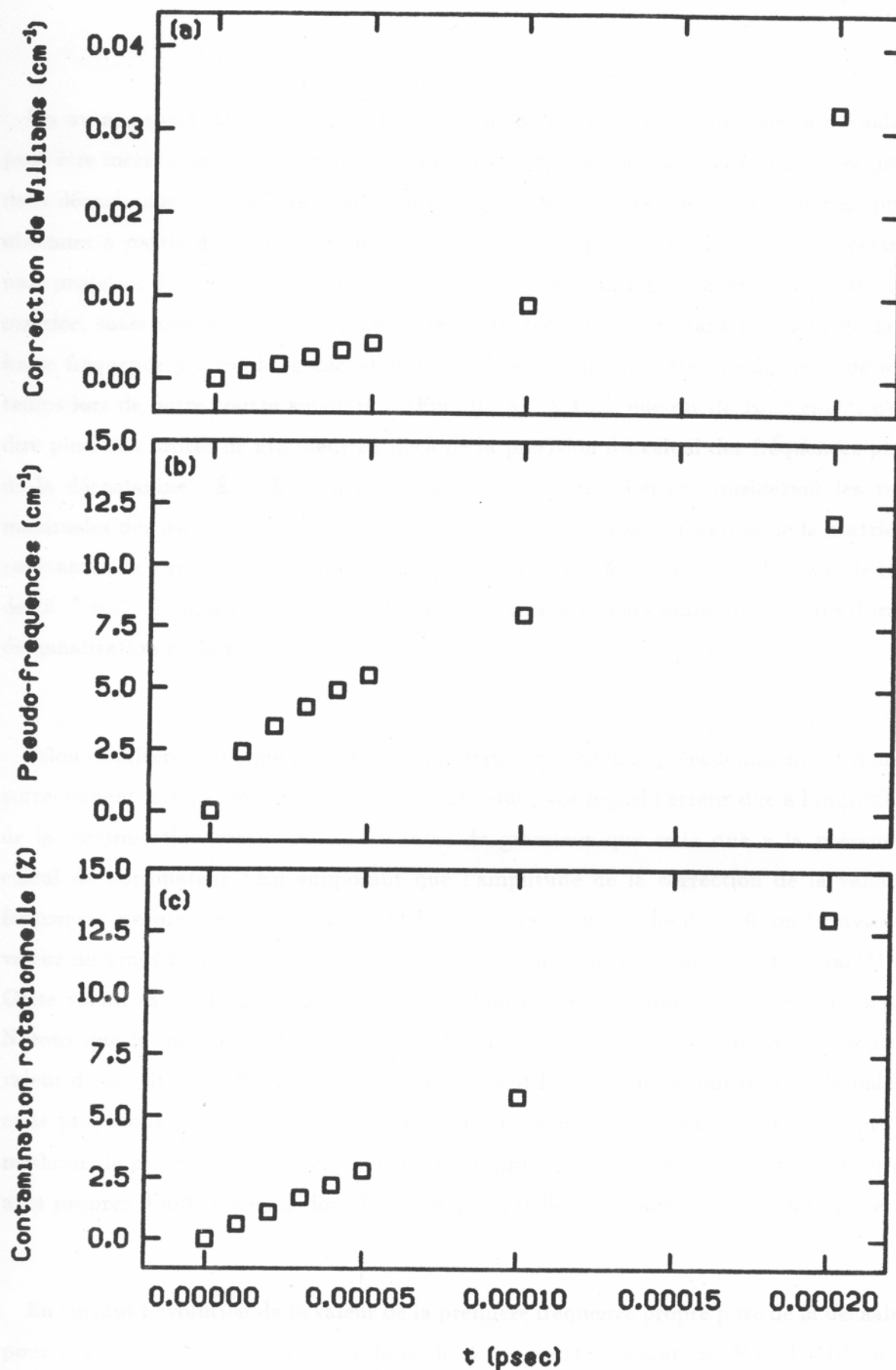


Figure II-2. Effet de la méthode de projection de Williams sur le calcul des modes de vibration de structures dont l'énergie potentielle est de plus en plus mal minimisée : (a) correction de Williams de la plus basse fréquence propre de la décaalanine; (b) somme des six pseudo-fréquences propres "brutes"; (c) contamination rotationnelle des axes propres "bruts".

Un autre aspect de la contamination rotationnelle des axes propres de la décaalanine peut être mesuré par l'amplitude de la correction affectant le calcul des fréquences propres de la décaalanine, c'est-à-dire la différence entre les valeurs des fréquences propres "pures", obtenues à partir de la matrice des constantes de force calculée dans l'espace cartésien, puis projetée, et celles des fréquences propres "brutes", obtenues à partir de cette même matrice, mais non projetée. L'amplitude de la correction affectant le calcul de la plus basse fréquence propre de la décaalanine n'évolue pas de manière linéaire en fonction du temps lors de notre courte simulation (Fig. II-2a). A $t=0$, elle est de 10^{-7} cm^{-1} , c'est-à-dire plusieurs ordres de grandeur en deçà de la précision du calcul des fréquences propres de la décaalanine. En effet, on peut estimer cette précision en considérant les valeurs maximales des pseudo-fréquences propres à la suite de la diagonalisation de la matrice des constantes de force projetée au préalable par la méthode de Williams : elles sont de l'ordre de 10^{-4} cm^{-1} (la non nullité de ces valeurs est due aux erreurs numériques faites durant la diagonalisation de la matrice).

Selon ce critère, l'énergie potentielle de la structure obtenue précédemment est donc très correctement minimisée. Un cas limite serait celui pour lequel l'erreur due à l'imperfection de la minimisation serait du même ordre de grandeur que celle due à la précision du calcul de l'ordinateur. En supposant que l'amplitude de la correction de la valeur des fréquences propres évolue linéairement lorsque t est très proche de $t=0$, on trouve que la valeur du gradient moyen correspondant à cette situation est d'environ $0,01 \text{ Kcal/Mole.Å}$. Cette valeur est donc une valeur supérieure limite pour le critère de fin de minimisation. Notons que la méthode de projection de Williams a permis de définir rigoureusement la valeur de ce critère de fin de minimisation. Il s'agit là d'une des applications originales que nous proposons pour cette méthode. L'étude ci-dessus montre cependant aussi que cette méthode de projection n'est pas nécessaire, en pratique, pour calculer les fréquences et les axes propres d'une molécule dont l'énergie potentielle a été suffisamment minimisée.

En suivant l'évolution de la valeur de la première fréquence propre pure de la décaalanine pour des structures prises tout au long de notre courte simulation (Fig. II-1b), on note

qu'après les $2 \cdot 10^{-4}$ premières psec, la valeur de cette fréquence est de $22,21 \text{ cm}^{-1}$ alors qu'elle est de $21,92 \text{ cm}^{-1}$ à $t=0$. Si l'on tient compte de cette source d'erreur dans le calcul des fréquences propres de la décaalanine, on constate qu'elle est négligeable, de l'ordre des erreurs de calcul de l'ordinateur, lorsque le gradient moyen vaut environ $0,0002 \text{ Kcal/Mole.Å}$. Pour tenir compte de cet effet, dont l'intensité est supérieure à l'erreur due à l'imperfection de la minimisation, il faut donc réviser à la baisse la valeur du critère de fin de minimisation obtenue ci-dessus. Notons que cet effet traduit le caractère anharmonique de la fonction énergie potentielle le long de cet axe propre, c'est-à-dire le fait que la courbure de la surface d'énergie potentielle n'a pas une valeur constante tout au long de l'axe du mode de vibration dont la fréquence est de $21,92 \text{ cm}^{-1}$. Ce résultat n'est pas, en lui-même, inattendu. Lors de tout changement de conformation, la courbure de la surface d'énergie potentielle doit changer petit à petit et devenir nulle, puis négative, dans une direction au moins de l'espace des configurations. Plus précisément, dans tout état de transition séparant deux conformations, par définition, les forces agissant sur les atomes sont nulles (la structure est en équilibre instable) et la courbure de la surface d'énergie potentielle est négative dans la direction du chemin suivi par la molécule lors de la transition. L'aspect instructif de notre résultat est le suivant : l'anharmonicité de la fonction énergie potentielle le long de cet axe propre se manifeste très rapidement, après $2 \cdot 10^{-4}$ psec, alors que les atomes se sont très peu déplacés, d'environ $0,0036 \text{ Å}$ chacun, en moyenne.

Cet effet anharmonique peut aussi être visualisé en suivant l'évolution en fonction du temps de la trace de la matrice des constantes de force projetée, $\mathbf{F_p}$, tout au long de notre courte simulation. Elle décroît linéairement (Fig. II-1c). Or, la trace de la matrice $\mathbf{F_p}$ est mathématiquement égale à la somme des valeurs propres de cette matrice, c'est-à-dire à la somme des carrés des pulsations des modes de vibration de la décaalanine. Globalement, la valeur des fréquences propres élevées de la décaalanine, qui dominent l'évolution de la valeur de la trace de $\mathbf{F_p}$, diminue donc. Cette diminution est relativement faible, 2 cm^{-1} au plus, entre le début et la fin de notre simulation, pour les fréquences dont la valeur avoisine 3200 cm^{-1} . En valeur absolue, cette diminution n'est pas beaucoup plus importante que la variation qui affecte la valeur des plus basses fréquences propres de la décaalanine mais, par contre, en valeur relative, elle est bien plus faible. Ce dernier

résultat traduit, et confirme, un résultat bien connu, à savoir que l'anharmonicité de la fonction énergie potentielle d'une molécule le long des axes de ses modes de vibration de basse fréquence est bien plus importante que le long des axes de ses modes de vibration de haute fréquence. Ce qui signifie ceci : la courbure de la surface d'énergie potentielle varie nettement moins, en valeur relative, le long des axes des modes de vibration de haute fréquence.

La précocité de l'apparition d'effets anharmoniques non négligeables le long des axes des modes de vibration de basse fréquence est confirmée par l'observation suivante, obtenue lors de l'analyse d'une trajectoire préalable : entre $t=2 \cdot 10^{-4}$ et $t=5 \cdot 10^{-4}$ psec, les premières fréquences propres imaginaires apparaissent. Ceci signifie que la surface d'énergie potentielle devient convexe dans certaines directions de l'espace des configurations. Par exemple, à $t=5 \cdot 10^{-4}$ psec, les fréquences correspondant à la courbure de la surface d'énergie potentielle le long de deux des axes propres de la décaalanine ont respectivement pour valeur $i23,21 \text{ cm}^{-1}$ et $i10,10 \text{ cm}^{-1}$. Corrélativement, la correction apportée par la projection de Williams aux valeurs de ces deux fréquences est importante, respectivement, $+2,64$ et $+7,10 \text{ cm}^{-1}$. On voit donc illustrée ici l'importance de la correction de Williams lorsqu'on cherche à déterminer les axes de l'espace des configurations le long desquels la surface d'énergie potentielle est convexe.

3) Fréquences propres de la décaalanine.

La structure de la décaalanine obtenue à l'issue de la minimisation POWELL vérifiant bien le critère de fin de minimisation défini ci-dessus, c'est en toute confiance que l'on peut utiliser les fréquences et les axes propres calculés à partir de cette configuration atomique puisque la précision de leur détermination numérique n'est limitée que par la précision des calculs de l'ordinateur. La figure II-3a montre les fréquences des modes de vibration de la décaalanine, disposées dans l'ordre dans lequel elles ont été calculées. Une première partie du spectre est constitué d'un quasi-continuum qui s'étale de $21,92 \text{ cm}^{-1}$ à $712,24$

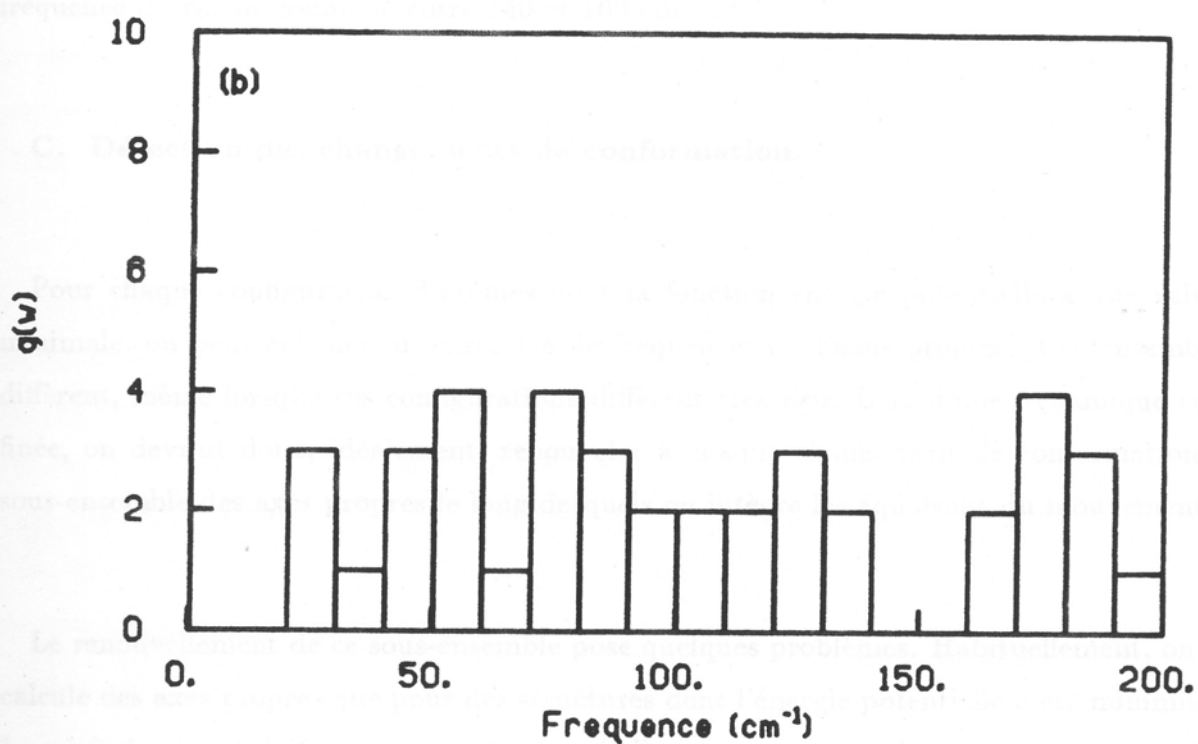
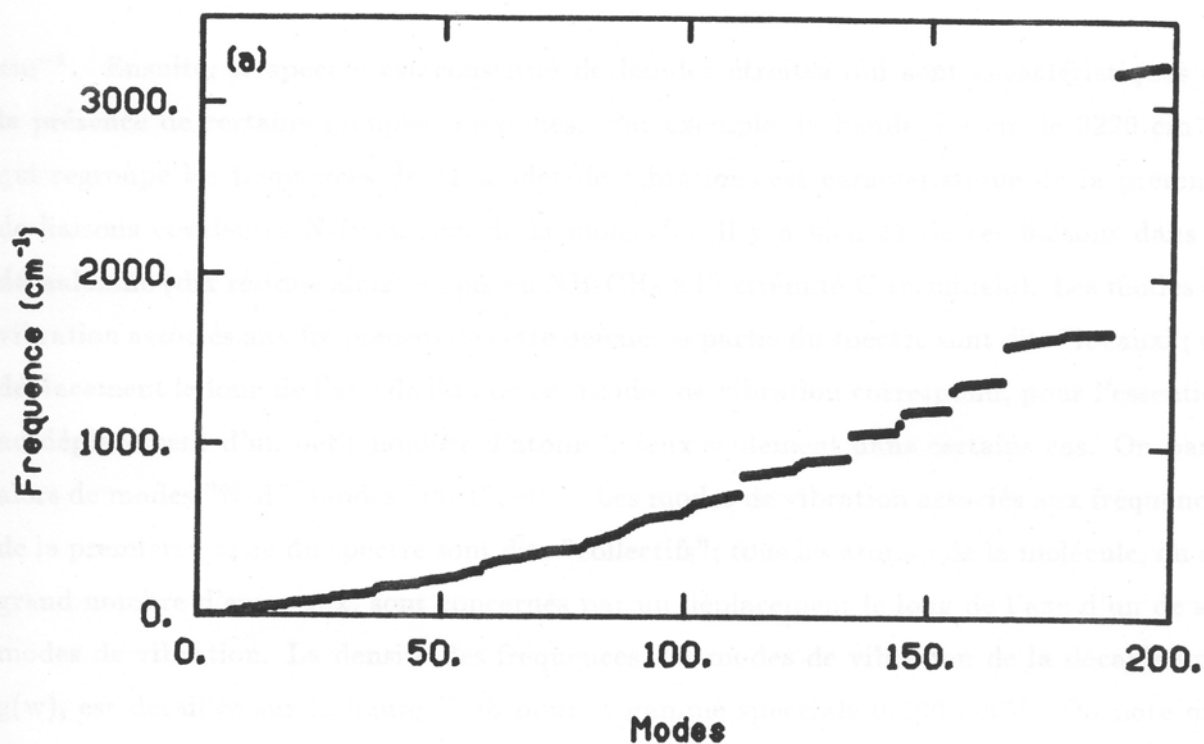


Figure II-3. Modes de vibration de la décaalanine : (a) fréquences des modes de vibration de la décaalanine, disposées dans l'ordre dans lequel elles sont calculées; (b) densité des fréquences des modes de vibration, $g(w)$, dans la gamme 0-200 cm^{-1} .

cm^{-1} . Ensuite, le spectre est constitué de bandes étroites qui sont caractéristiques de la présence de certains groupes chimiques. Par exemple, la bande autour de 3220 cm^{-1} , qui regroupe les fréquences de 11 modes de vibration, est caractéristique de la présence de liaisons covalentes N-H au sein de la molécule. Il y a bien 11 de ces liaisons dans la décaalanine (dix résidus alanine plus un NH-CH_3 à l'extrémité C-terminale). Les modes de vibration associés aux fréquences de cette deuxième partie du spectre sont dits "locaux"; un déplacement le long de l'axe de l'un de ces modes de vibration correspond, pour l'essentiel, au déplacement d'un petit nombre d'atomes, deux seulement dans certains cas. On parle alors de modes "N-H", modes "O-H", etc... Les modes de vibration associés aux fréquences de la première partie du spectre sont dits "collectifs"; tous les atomes de la molécule, ou un grand nombre d'entre eux, sont concernés par un déplacement le long de l'axe d'un de ses modes de vibration. La densité des fréquences des modes de vibration de la décaalanine, $g(w)$, est détaillée sur la figure II-3b pour la gamme spectrale $0-200 \text{ cm}^{-1}$. On note que cette distribution de fréquences n'est ni régulière ni continue. Par exemple, il n'y a aucune fréquence de valeur comprise entre 140 et 160 cm^{-1} .

C. Détection des changements de conformation.

Pour chaque configuration d'atomes dont la fonction énergie potentielle a une valeur minimale, on peut calculer un ensemble de fréquences et d'axes propres. Ces ensembles diffèrent, même lorsque ces configurations diffèrent très peu. Lors d'une dynamique confinée, on devrait donc, idéalement, renouveler à chaque changement de conformation le sous-ensemble des axes propres le long desquels on intègre les équations du mouvement.

Le renouvellement de ce sous-ensemble pose quelques problèmes. Habituellement, on ne calcule des axes propres que pour des structures dont l'énergie potentielle a été minimisée. La minimisation de la fonction énergie potentielle est un processus long qui nécessite parfois plusieurs milliers d'étapes, c'est-à-dire plusieurs milliers de calculs de l'énergie potentielle de la molécule. Or, le facteur temps est crucial puisque l'un des principaux objectifs de la

méthode de dynamique confinée est de pouvoir en gagner lors du calcul d'une trajectoire de dynamique moléculaire. Si on ne minimise pas bien la fonction énergie potentielle d'une molécule, on obtient, comme on l'a vu dans la partie B2 de ce chapitre, quelques axes propres le long desquels la surface d'énergie potentielle est convexe. Il n'est pas alors possible de savoir, à priori, si les mouvements qui ont lieu le long de ces axes sont rapides ou lents, d'ordinaire. Or, les équations du mouvement des atomes ne peuvent être résolues avec un grand pas d'intégration que le long des axes des modes de vibration de basse fréquence. En d'autres termes, la méthode de dynamique confinée ne peut être par principe efficace que si l'on peut savoir quels sont les axes de l'espace des configurations le long desquels la courbure de la surface d'énergie potentielle est faible tout au long de la simulation. Or, il n'y a aucun moyen pour inférer la valeur de la courbure de la surface d'énergie potentielle au fond d'un puits de potentiel à partir de la mesure de la courbure de cette surface en un point de la paroi de ce puits. Si en un point de la surface de la paroi d'un puits de potentiel la courbure est nulle, la courbure au fond de ce même puits peut aussi bien être faible que très importante. Ecarter les axes propres le long desquels la surface d'énergie potentielle est convexe ne résout pas le problème car ces axes sont physiquement importants, puisqu'une molécule longe forcément de tels axes lorsqu'elle change de conformation.

Par ailleurs, si l'on renouvelait au cours d'une simulation les axes propres le long desquels on intègre les équations du mouvement, il faudrait, pour assurer la conservation de l'énergie totale, transférer sur les "nouveaux" axes propres la totalité de l'énergie cinétique présente sur les "anciens". Comme les deux sous-espaces définis par ces deux ensembles d'axes sont différents, il faudrait, pour maintenir l'énergie totale constante, ajuster la valeur de l'énergie cinétique à sa valeur précédente, ce qui introduirait une rupture dans le comportement dynamique de la molécule.

D'un point de vue plus pratique, la comparaison directe de trajectoires au cours desquelles se produisent des changements de conformation n'est pas évidente. Or, pour tester la validité de notre méthode de dynamique moléculaire confinée, nous devons com-

parer les trajectoires obtenues par cette méthode aux trajectoires obtenues par les méthodes traditionnelles. Nous allons donc limiter notre étude à celle des propriétés intrinsèques de la méthode de dynamique confinée, en étudiant des trajectoires au cours desquelles ne se produit aucun changement de conformation. Par ailleurs, comme l'efficacité espérée de la méthode de dynamique confinée repose sur le caractère hautement harmonique du comportement dynamique des molécules biologiques, cette étude sera celle de la validité de l'approximation harmonique lorsque la molécule évolue dans le fond d'un seul puits de la surface d'énergie potentielle, c'est-à-dire dans le meilleur des cas possibles. En général, lorsqu'on étudie les aspects "anharmoniques" de la dynamique d'une molécule, les effets que l'on observe sont dûs pour une part au fait que la molécule passe d'un puits à un autre de la surface d'énergie potentielle, et pour une autre part au fait que la forme de la surface d'énergie potentielle n'est parabolique que très près du fond de ces puits, sans qu'il soit possible de quantifier les rôles respectifs de ces deux types de causes. Nous avons donc l'occasion ici d'étudier spécifiquement l'une de ces deux formes d'anharmonicité.

Comme la décaalanine est une petite molécule stable, des trajectoires de durée significative, sans aucun changement de conformation, fut-ce très localisé, sont possibles. Par ailleurs, comme la conformation de cette molécule est aisée à décrire à l'aide de quelques paramètres, puisqu'il s'agit d'une hélice α , il est aussi aisé de détecter le moindre petit changement de conformation. Pour ce faire, on calcule à chaque étape d'une dynamique les caractéristiques structurales de l'hélice, c'est-à-dire son pas, son rayon, et l'angle moyen que font les plans des liaisons peptidiques de deux résidus successifs. Par ailleurs, on suit l'évolution temporelle de la distance moyenne, $d(t)$, entre la structure de la décaalanine à chaque instant de la simulation et sa structure initiale :

$$d(t) = \sqrt{\frac{\sum_{i=1}^{3N} (r_i(t) - r_i(0))^2}{N}}$$

$r_i(t)$ étant la coordonnée cartésienne i de la molécule à l'instant t . On étudie ensuite les structures extraites de la trajectoire pour lesquelles ces paramètres ont des valeurs extrêmes. Pour cela, l'énergie potentielle de chacune de ces structures est minimisée. Si la valeur de l'énergie potentielle d'une des structures obtenues à l'issue de ces minimisations

ou si l'une des valeurs des angles dièdres de ces structures diffère significativement de celles de la structure initiale, on conclut qu'un changement de conformation a eu lieu et la trajectoire n'est pas retenue pour les analyses ultérieures.

Il n'y a aucun changement de conformation durant les 10 premières picosecondes d'une trajectoire de la décaalanine à 300°K, calculée par la méthode traditionnelle de dynamique moléculaire. Or, durant 10 psec, une quinzaine d'oscillations doivent se produire le long de l'axe du mode de vibration dont la fréquence est la plus basse ($21,9 \text{ cm}^{-1}$), ce qui est suffisant pour obtenir des valeurs moyennes statistiquement satisfaisantes. Cette trajectoire de 10 psec est la trajectoire de référence à laquelle les trajectoires de dynamique confinée sont comparées dans le présent travail. Notons que toutes ces trajectoires ont les mêmes conditions initiales. Elles ne diffèrent que par les contraintes qui sont imposées aux déplacements des atomes.

D. Trajectoires de dynamiques confinées.

Une trajectoire de dynamique moléculaire est considérée comme correcte, d'un point de vue statistique, lorsque les fluctuations de l'énergie totale, qui devraient en toute rigueur être nulles, sont, au cours de la simulation, dix fois plus faibles que les fluctuations de l'énergie cinétique (H.J.C.Berendsen & W.F.Van Gunsteren, 1984). La valeur du rapport entre ces deux types de fluctuations permet de comparer des trajectoires calculées avec des méthodes ou des pas d'intégration différents. La figure II-4 montre, pour différentes méthodes de simulation, l'évolution de ce rapport en fonction du pas temporel utilisé pour intégrer les équations du mouvement des atomes de la décaalanine. Toutes les simulations ainsi comparées durent 10 psec. Deux méthodes de simulation servent de référence, la méthode traditionnelle proprement dite et une méthode traditionnelle incluant une étape supplémentaire, durant laquelle les longueurs de toutes les liaisons interatomiques sont contraintes au moyen de l'algorithme SHAKE. Les autres méthodes de simulation étudiées sont toutes des méthodes de dynamique moléculaire confinée. Ces méthodes diffèrent

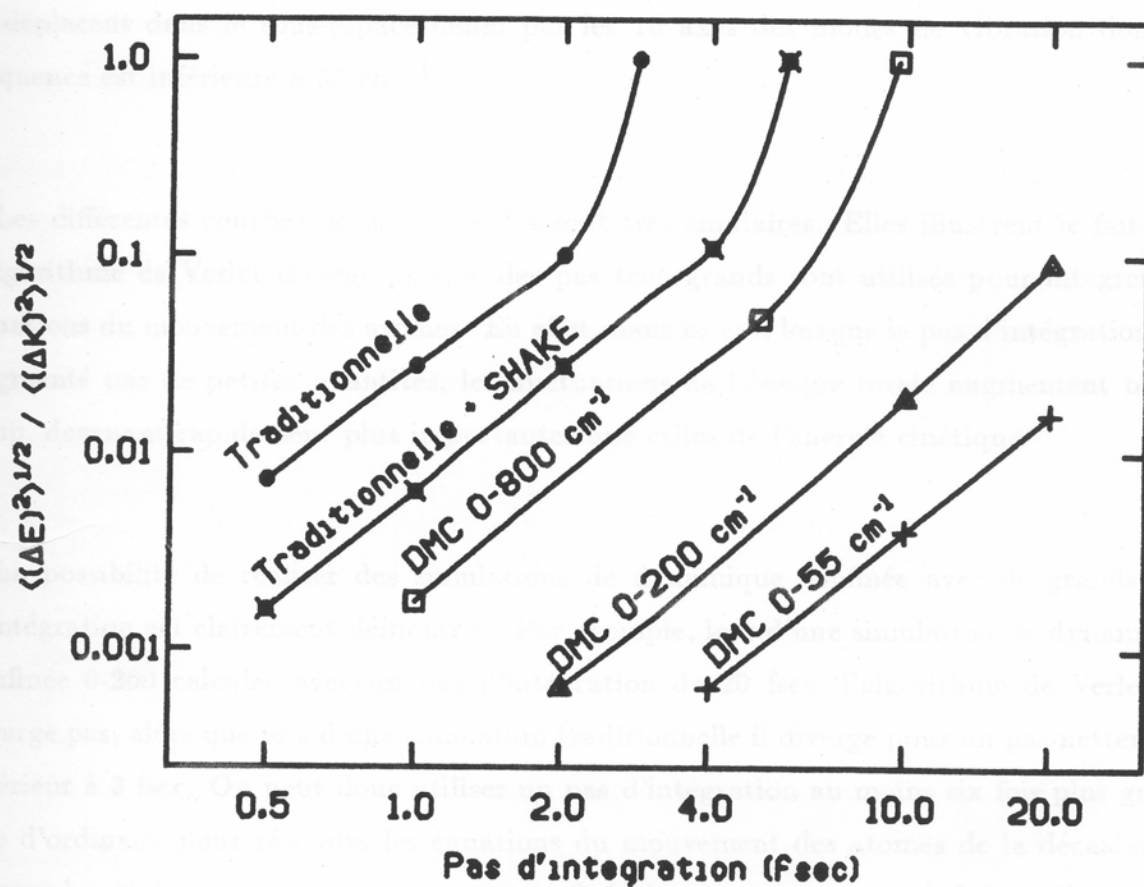


Figure II-4. Etude de la validité statistique de trajectoires de dynamique moléculaire traditionnelles et confinées de la décaalanine : rapport des fluctuations de l'énergie totale et des fluctuations de l'énergie cinétique en fonction de la valeur du pas d'intégration utilisé, pour différents types de simulation.

par le choix du sous-espace des configurations dans lequel les atomes sont libres de se mouvoir. Par exemple, lors d'une dynamique moléculaire confinée (DMC) 0-55, les atomes se déplacent dans le sous-espace défini par les 16 axes des modes de vibration dont la fréquence est inférieure à 55 cm^{-1} .

Tableau II-2 : Gain de temps CPU obtenu lors de l'utilisation de la méthode de Verlet.

Les différentes courbes de la figure II-4 sont très similaires. Elles illustrent le fait que l'algorithme de Verlet diverge lorsque des pas trop grands sont utilisés pour intégrer les équations du mouvement des atomes. En effet, dans ce cas, lorsque le pas d'intégration est augmenté par de petites quantités, les fluctuations de l'énergie totale augmentent beaucoup, devenant rapidement plus importantes que celles de l'énergie cinétique.

Tableau II-3 : Comparaison des temps CPU pour la dynamique confinée et traditionnelle.

La possibilité de réaliser des simulations de dynamique confinée avec de grands pas d'intégration est clairement démontrée. Par exemple, lors d'une simulation de dynamique confinée 0-200 calculée avec un pas d'intégration de 20 fsec, l'algorithme de Verlet ne diverge pas, alors que lors d'une simulation traditionnelle il diverge pour un pas nettement inférieur à 3 fsec. On peut donc utiliser un pas d'intégration au moins six fois plus grand que d'ordinaire pour résoudre les équations du mouvement des atomes de la décaalanine lorsque les déplacements des atomes sont confinés dans un sous-espace défini par l'ensemble des axes des modes de vibration dont la fréquence est inférieure à 200 cm^{-1} . Comme la méthode de confinement employée ci-dessus ne rallonge que de 10% la durée du calcul d'un pas de dynamique, il faut donc environ six fois moins de temps CPU pour calculer 10 psec d'une trajectoire de dynamique confinée 0-200 que pour calculer une trajectoire de même durée par une méthode traditionnelle.

Tableau II-4 : Comparaison des temps CPU pour la dynamique confinée et traditionnelle.

Afin de quantifier plus précisément le gain de temps CPU que permettent les différents types de dynamique confinée, les pas d'intégration pour lesquels le rapport des fluctuations de l'énergie totale et de l'énergie cinétique a la valeur de 0,01 sont comparés (tableau II-2). On constate tout d'abord que l'algorithme SHAKE permet d'utiliser un pas d'intégration deux fois plus grand que lors d'une dynamique traditionnelle. C'est ce qui est couramment

admis. La méthode de comparaison ci-dessus est donc pertinente. On constate ensuite qu'une dynamique confinée 0-55 peut être calculée avec un pas presque 30 fois plus grand que de coutume, ce qui est considérable.

Tableau II-2 : Gain de temps CPU optimal dû à l'utilisation de la méthode de dynamique moléculaire confinée. Ce gain est mesuré en comparant, pour diverses méthodes de simulation, les pas temporel d'intégration pour lesquels le rapport des fluctuations de l'énergie totale à celles de l'énergie cinétique est de 0,01 (ces valeurs sont tirées de la figure II-4). Ce gain est comparé à sa valeur idéale, c'est-à-dire à ce qu'il serait si la décaalanine avait un comportement dynamique parfaitement harmonique.

Dynamique	Pas	Pas idéal	Gain	Gain idéal
Traditionnelle	0,6	0,6	1	1
Idem + SHAKE	1,2	...	2	...
DMC 0-800	2,4	2,7	4	5
DMC 0-200	7,6	9,8	13	16
DMC 0-55	16,1	35,4	27	59

La méthode de dynamique confinée permet donc bien d'utiliser des pas d'intégration importants pour résoudre les équations du mouvement des atomes. Néanmoins, si la décaalanine était un système harmonique idéal, l'augmentation du pas d'intégration serait encore plus importante. En effet, c'est l'intégration du mouvement le plus rapide possible lors d'une simulation qui limite la valeur du pas d'intégration car, pour une précision d'intégration donnée, il faut un nombre de pas constant pour intégrer une période d'un mouvement vibratoire. Comme la fréquence propre la plus élevée de la décaalanine lors d'une dynamique confinée 0-55 est de $53,8 \text{ cm}^{-1}$, et comme la fréquence propre la plus élevée de la décaalanine est, en absence de toute contrainte, de $3241,8 \text{ cm}^{-1}$; si la décaalanine était un système harmonique idéal, le pas utilisé lors de la simulation de dynamique confinée 0-55 pourrait être environ 60 fois plus important que lors d'une simulation de dynamique traditionnelle, c'est-à-dire de 35,4 psec au lieu de 0,6 psec, dans le cas où l'on souhaite intégrer les équations du mouvement des atomes avec une très grande précision (voir tableau II-2). La différence entre cette valeur idéale et la valeur observée est

une manifestation de l'anharmonicité de la fonction énergie potentielle le long des axes des modes de vibration de fréquence inférieure à 55 cm^{-1} . Cette différence croît régulièrement lorsqu'on utilise les axes de modes de vibration dont les fréquences sont de plus en plus basses (tableau II-2), ce qui illustre une nouvelle fois le fait que le caractère anharmonique des mouvements des atomes d'une molécule est plus marqué le long des axes des modes de vibration de basse fréquence.

E. Comparaison des méthodes de dynamique moléculaire confinée et traditionnelle.

La méthode de dynamique confinée permet d'intégrer les équations du mouvement des atomes avec un pas temporel important. Cependant, cette méthode n'est intéressante que si les trajectoires ainsi calculées sont peu différentes des trajectoires obtenues par les méthodes traditionnelles, à partir de conditions initiales identiques. Une trajectoire calculée par une méthode de dynamique confinée $0-\nu$ serait néanmoins très utile si l'évolution des projections sur les P axes du sous-espace dans lequel les atomes se déplacent était similaire à celle des projections sur ces mêmes P axes lors d'une trajectoire obtenue par une méthode traditionnelle. Pour vérifier cela, nous utilisons une méthode originale, dite de dynamique moléculaire projetée, dont le principe est analogue à celui de la méthode de dynamique moléculaire confinée.

Une trajectoire de dynamique moléculaire projetée $0-\nu$ résulte des déplacements des atomes, lors de la dynamique traditionnelle, dans le sous-espace dans lequel les atomes se déplacent au cours de la dynamique confinée $0-\nu$. Pour calculer cette trajectoire, l'évolution des coordonnées propres, $\mathbf{q}^P(\mathbf{t})$, sur les axes des P modes de vibration dont la fréquence est inférieure à ν est déduite de celle des coordonnées pondérées des atomes lors de la dynamique traditionnelle, $\mathbf{r}(\mathbf{t})$, par la relation suivante, obtenue à partir de l'équation (13):

$$\mathbf{q}^P(\mathbf{t}) = \mathbf{A}_P^t \mathbf{r}(\mathbf{t})$$

où A_P^t est la sous-matrice de A^t de dimensions $P \times 3N$. Ensuite, on calcule la trajectoire issue de l'évolution des coordonnées propres sur ces seuls P axes propres :

$$r^P(t) = A_P q^P(t)$$

et l'opération globale effectuée est donc :

$$r^P(t) = A_P A_P^t r(t) \quad (42)$$

Il s'agit de la même opération de projection que celle qui résume la méthode de dynamique moléculaire confinée (équation 40). La différence entre une trajectoire de dynamique projetée et une trajectoire de dynamique confinée peut donc se décrire ainsi : dans la première, la projection est effectuée sur une trajectoire calculée auparavant, tandis que dans la seconde, la projection est effectuée lors du calcul de la trajectoire. Dans le second cas, la trajectoire est perturbée par l'opération de projection, puisque les déplacements des atomes sont soumis à des contraintes, tandis que dans le premier cas elle ne l'est pas. Dans les deux cas, les atomes se déplacent dans le même sous-espace défini par les mêmes axes. Les trajectoires de dynamique projetée $0-\nu$, issues d'une trajectoire de dynamique traditionnelle via l'équation (42), sont donc les trajectoires de référence idéales pour analyser les effets du confinement sur le déroulement d'une simulation.

La figure II-5 montre les fluctuations moyennes de la position des atomes C_α pour différentes trajectoires de dynamique projetée $0-\nu$ et confinée $0-\nu$. Il s'agit de de trajectoires de 10 psec de la décaalanine à 300°K. On a calculé, avec un pas temporel d'intégration de 0,001 psec et des conditions initiales identiques, 7 trajectoires de dynamique confinée, respectivement de type 0-0, 0-20, 0-50, 0-200, 0-800, 0-1800, 0-3250, et une trajectoire de dynamique traditionnelle de laquelle on a tiré 7 trajectoires de dynamique projetée comparables aux 7 trajectoires de dynamique confinée. Les axes des 6 modes de vibration de fréquence 0-0 sont les axes des pseudo-modes de rotation et de translation. Comme la décaalanine ne tourne pas durant nos simulations, les atomes C_α ne se déplacent pas dans les directions qui font globalement tourner ou translater la molécule et les fluctuations de la position des atomes C_α sont pratiquement nulles aussi bien lors d'une dynamique confinée 0-0 que lors d'une dynamique projetée 0-0.

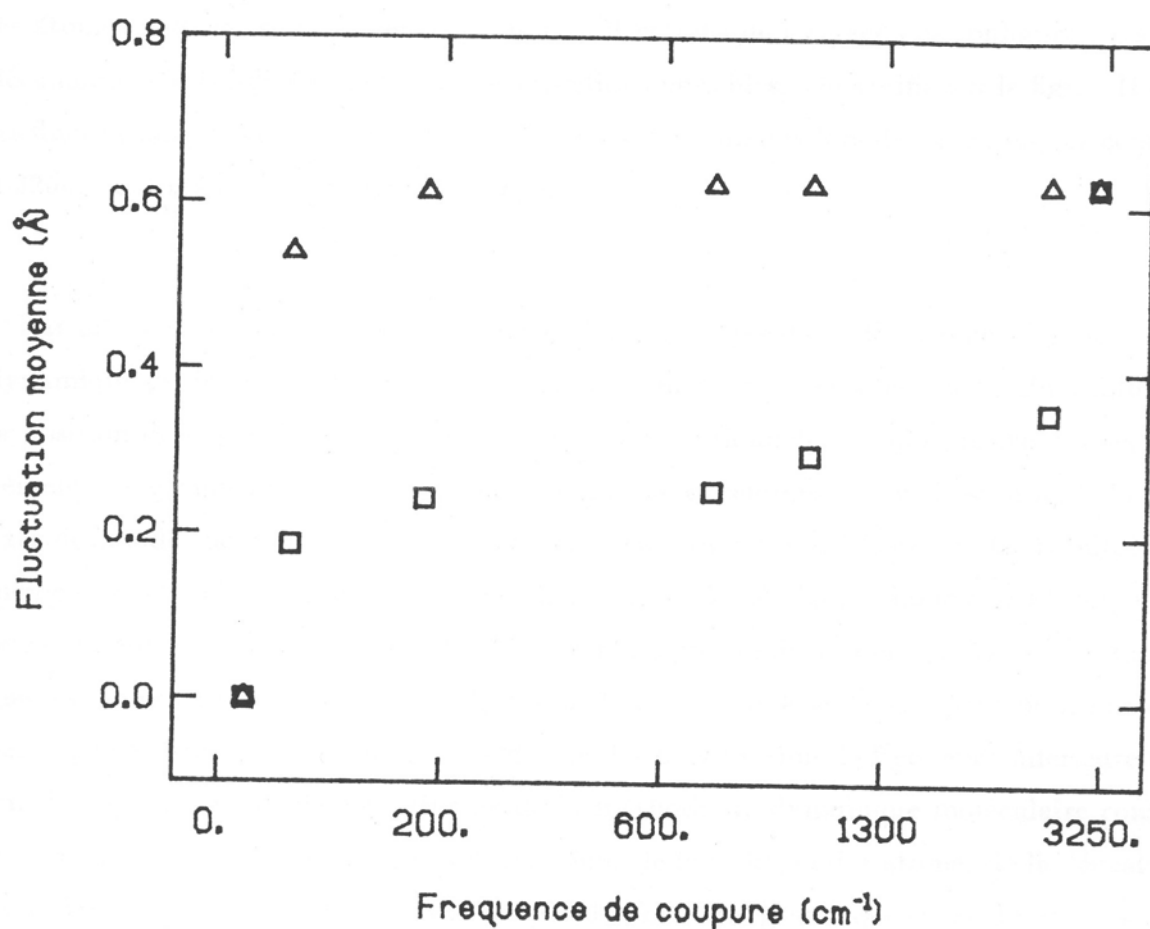


Figure II-5. Comparaison de trajectoires de la décaalanine à 300°K, obtenues par des méthodes de dynamique moléculaire confinée ou traditionnelle : fluctuations de la position des atomes C_{α} lors de dynamiques moléculaires confinées (carrés) ou projetées (triangles) dans le sous-espace défini par les axes des modes de vibration de fréquence inférieure à la fréquence de coupure.

Les axes des modes de vibration de fréquence comprises entre 0 et 3250 cm^{-1} regroupent l'ensemble des 198 axes propres de la décaalanine. Lors d'une dynamique confinée 0-3250 les atomes peuvent se déplacer dans les 198 directions de l'espace des configurations de la décaalanine, c'est-à-dire dans toutes les directions possibles. On vérifie sur la figure II-5 que les fluctuations de la position des atomes C_α sont similaires lors des dynamiques confinées 0-3250, projetée 0-3250, et traditionnelle.

Par ailleurs, on constate que les fluctuations de la position des atomes C_α lors d'une dynamique projetée 0-200 sont de 5 % seulement moins importantes que les fluctuations de la position de ces atomes lors de la dynamique traditionnelle, ce qui prouve que les mouvements de grande amplitude de la décaalanine se développent pour l'essentiel le long des axes des modes de vibration dont la fréquence est inférieure à 200 cm^{-1} . Ce résultat est le même que celui obtenu pour un système harmonique idéal. La méthode de dynamique projetée, parente naturelle de la méthode de dynamique confinée, permet donc de démontrer que les mouvements de grande amplitude de la décaalanine se développent pour l'essentiel dans le sous-espace défini par les axes des modes de vibration de fréquence inférieure à 200 cm^{-1} . Ce résultat justifie le principe de la méthode de dynamique moléculaire confinée. Pouvoir rendre compte de 95% des fluctuations de la position des atomes de la décaalanine avec des trajectoires calculées avec un pas 13 fois plus grand que celui d'une trajectoire traditionnelle, c'est-à-dire environ 13 fois plus rapidement, serait déjà un excellent résultat pour la méthode de dynamique moléculaire confinée, 6 fois au delà des performances de la méthode basée sur l'algorithme SHAKE, par exemple.

Cependant, en comparant les trajectoires de dynamique projetée 0-200 et de dynamique confinée 0-200, on constate que les fluctuations de la position des atomes C_α lors de cette dernière dynamique sont de plus de 50% inférieures aux fluctuations de la position des atomes C_α lors de la première. Notre méthode de confinement des déplacements des atomes dans le sous-espace défini par les axes des modes de vibration de fréquence inférieure à 200 cm^{-1} réduit donc de manière très importante les fluctuations de la position des atomes C_α . De plus, ces fluctuations n'augmentent pas significativement lorsque ce

sous-espace est agrandi en ajoutant aux précédents les axes des modes de vibration dont la fréquence est comprise entre 200 et 800 cm^{-1} . Lors d'une dynamique confinée 0-800, les fluctuations de la position des atomes C_α sont d'environ 50% moindres que celles lors d'une dynamique traditionnelle. L'augmentation la plus importante de ces fluctuations est obtenue lorsqu'on agrandit le sous-espace dans lequel se déplacent les atomes en rajoutant les axes des modes de vibration dont la fréquence est comprise entre 800 et 930 cm^{-1} , puis les axes des modes de vibration dont la fréquence est comprise entre 3200 et 3250 cm^{-1} . Ce dernier groupe de fréquences correspond aux directions des modes de vibration "N-H".

La méthode de dynamique moléculaire projetée a montré que les déplacements de la décaalanine le long des axes des modes de vibration de haute fréquence sont négligeables, lors d'une dynamique traditionnelle. Cependant, la méthode de dynamique moléculaire confinée a montré que les déplacements le long de ces axes sont nécessaires pour reproduire les résultats des dynamiques moléculaires traditionnelles. En d'autres termes, il existe un couplage entre les mouvements se produisant le long des axes des modes de vibration de haute fréquence et ceux se produisant le long des axes des modes de vibration de basse fréquence. Cette interaction, ce couplage, est une manifestation directe de l'anharmonicité de la fonction énergie potentielle de la décaalanine. En effet, les déplacements d'un système harmonique idéal le long d'un axe propre donné ne perturbent en aucune façon les déplacements du système le long de ses autres axes propres. Il sont totalement décorrélés. Cette propriété est une des caractéristiques majeures des systèmes dont le comportement est harmonique. En termes de couplage, la figure II-5 suggère que les déplacements de la décaalanine le long des axes de basse fréquence, 0-200 par exemple, sont couplés à ceux se produisant, d'une part le long des axes 800-950, d'autre part le long des axes 3000-3250, au moins en ce qui concerne certains d'entre eux. La comparaison des fluctuations de la position des atomes C_α lors de dynamiques projetées obtenues à partir de trajectoires de dynamique confinée 0-800, 0-950 et 0-1800 (courbes non montrées) renforce cette hypothèse.

Afin de confirmer que ces couplages ont bien pour origine l'anharmonicité de la fonction

énergie potentielle de la décaalanine, et ne sont pas inhérentes à la méthode de confinement même, on a réalisé exactement la même étude que ci-dessus, mais à partir de trajectoires de la décaalanine à 5°K, cette fois. A cette basse température, les déplacements des atomes sont moins amples qu'à 300°K et les approximations sur lesquelles repose la théorie des modes normaux ne peuvent qu'être mieux vérifiées. La figure II-6 est, pour ces dynamiques à basse température, l'équivalent de la figure II-5. On constate cette fois que les fluctuations de la position des atomes C_α lors d'une dynamique confinée 0-200 sont de 2% inférieures aux fluctuations de la position de ces atomes lors d'une dynamique projetée 0-200. En d'autres termes, les couplages entre les mouvements des atomes le long des axes des modes de vibration de haute et de basse fréquence sont pratiquement négligeables à 5°K. Le phénomène qu'illustre la figure II-5 est donc bien une manifestation de l'anharmonicité de la fonction énergie potentielle de la décaalanine.

Afin d'avoir une première idée de l'origine des couplages mis en évidence par la méthode de dynamique moléculaire confinée à température ordinaire, les évolutions de la coordonnée propre sur l'axe du mode de vibration de plus basse fréquence lors des dynamiques projetée 0-55 et confinée 0-55 sont comparées (figure II-7). Si la décaalanine avait un comportement dynamique parfaitement harmonique, les courbes montrées seraient des sinusoides et leur transformée de Fourier ne donnerait qu'une seule fréquence, la fréquence de ce mode de vibration ($21,9 \text{ cm}^{-1}$). On observe bien cela lorsqu'on analyse des trajectoires de la décaalanine à 5°K. Par contre, pour des trajectoires de la décaalanine à 300°K, ce n'est pas le cas. La courbe obtenue n'est pas une simple sinusoïde (Fig. II-7a). Cependant, la fréquence moyenne du mouvement de la décaalanine le long de cet axe au cours de la trajectoire de dynamique traditionnelle est très proche de la valeur théorique. Lors d'une dynamique confinée 0-55, non seulement la trajectoire le long de l'axe du mode de vibration de plus basse fréquence n'est pas sinusoïdale, mais la fréquence moyenne dominant le spectre de cette trajectoire est bien supérieure à la fréquence de ce mode de vibration (Fig. II-7b). Ce résultat est généralisable à l'ensemble des axes le long desquels la décaalanine se déplace lors d'une dynamique confinée 0-55. Le décalage de la fréquence moyenne des mouvements de la décaalanine le long de ces axes vers les hautes fréquences est donc l'expression de la contrainte que notre méthode impose aux déplacements des atomes.

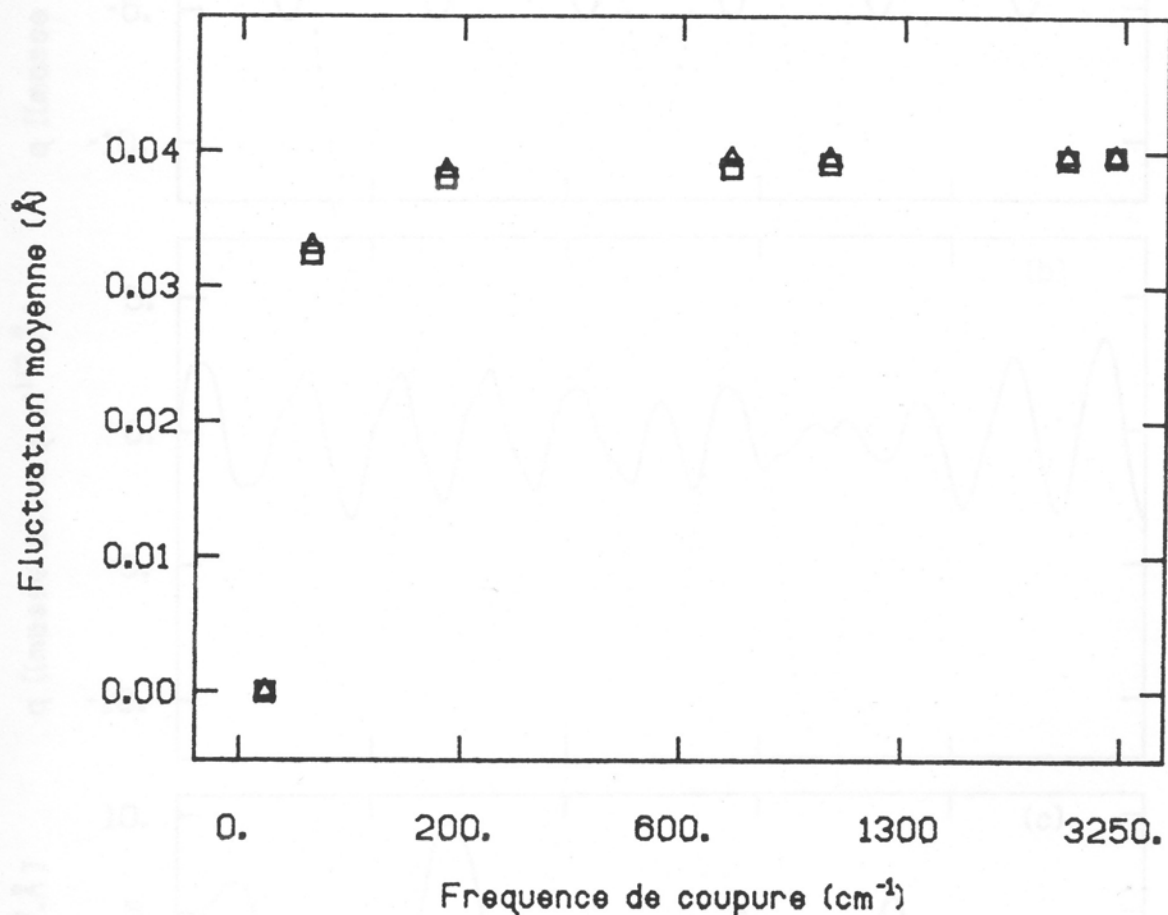


Figure II-6. Comparaison de trajectoires de la décaalanine à 5°K, obtenues par des méthodes de dynamique moléculaire confinée ou traditionnelle : fluctuations de la position des atomes C α lors de dynamiques moléculaires confinées (carrés) ou projetées (triangles) dans le sous-espace défini par les axes des modes de vibration de fréquence inférieure à la fréquence de coupure.

La théorie des modes normaux appliquée à des systèmes partiellement contraints met en évidence de tels décalages (voir, par exemple, Tobias et Ch.Brooks III, 1989, dans le cas de contraintes appliquées à des coordonnées internes quelconques). Or, lorsque la fréquence d'un oscillateur harmonique augmente, à énergie totale donnée, l'amplitude de son mouvement diminue proportionnellement. De fortes contraintes imposées aux déplacements des atomes, puisqu'elles augmentent la fréquence moyenne de leurs mouvements, doivent donc aussi réduire en moyenne leur amplitude. La figure II-7b illustre bien comment naissent les contraintes qu'impose la méthode de dynamique confinée. Pendant environ 0,1 picoseconde, les trajectoires issues des dynamiques confinées et traditionnelles évoluent de concert. Puis, elles divergent. Tout se passe comme si apparaissaient des forces qui s'opposent à un déplacement ample des atomes le long des axes des modes de vibration de basse fréquence lorsque la dynamique est confinée.

Afin d'élucider la nature de ces contraintes, on étudie plus particulièrement la dynamique traditionnelle réalisée avec l'algorithme SHAKE appliqué à l'ensemble des longueurs des liaisons covalentes. En effet, lors d'une telle dynamique, 65 degrés de liberté du système sont contraints, puisqu'il y a 65 liaisons covalentes dans la décaalanine. Or, les fluctuations de la position des atomes sont peu affectées. Si l'on considère la trajectoire de la décaalanine le long de l'axe du mode de vibration de plus basse fréquence (figure II-7c), lors de cette simulation, celle-ci est vraiment très proche de la trajectoire de la décaalanine le long de cet axe, lors de la simulation traditionnelle. Des contraintes exercées sur les longueurs des liaisons covalentes perturbent donc peu les mouvements de la décaalanine le long des axes de ses modes de vibration de basse fréquence.

Le couplage observé n'est donc pas le résultat de l'interaction directe des mouvements de haute fréquence avec les mouvements de basse fréquence, puisque dans la simulation étudiée ci-dessus les premiers n'ont pas lieu. Le couplage est donc de nature géométrique. Pour confirmer et préciser ce résultat, et afin de raisonner sur un cas simplifié, les trajectoires de la décaalanine le long des axes des modes de vibration "N-H" sont plus particulièrement considérées. Lorsque les atomes ne peuvent se déplacer dans le sous-espace défini par ces

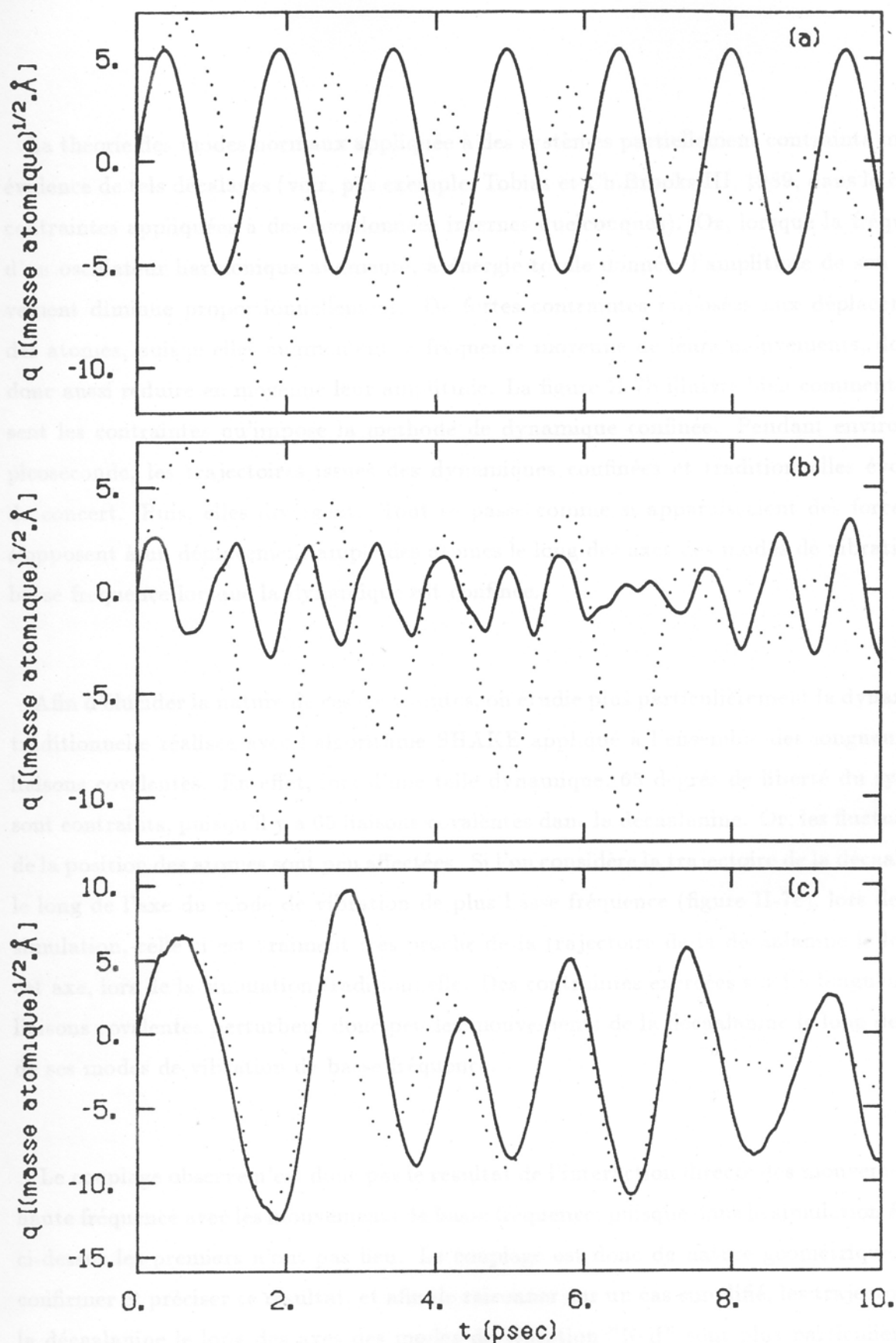


Figure II-7. Projection des déplacements des atomes de la décaalanine à 300°K sur l'axe du mode de vibration de fréquence 21.9 cm^{-1} : a) théorie des modes normaux; b) dynamique moléculaire confinée 0-55; c) dynamique moléculaire traditionnelle durant laquelle les longueurs de toutes les liaisons sont maintenues constantes au moyen de l'algorithme SHAKE. A des fins de comparaison, la trajectoire correspondante obtenue lors d'une simulation de dynamique moléculaire traditionnelle est représentée en pointillé sur chacune des trois figures.

axes, c'est-à-dire lors d'une dynamique confinée 0-1800, les fluctuations de la position des atomes C_α sont réduites de plus de 30% par rapport à ce qu'elles sont lors d'une trajectoire de dynamique traditionnelle. Or, un déplacement de la décaalanine le long d'un de ces axes correspond pour l'essentiel à une variation de la longueur d'une de ses liaisons N-H, lorsque sa configuration atomique est proche de sa configuration initiale.

La figure II-8c montre l'évolution en fonction du temps de la coordonnée propre sur l'axe du mode de vibration "N-H" de plus haute fréquence lors d'une dynamique traditionnelle, lorsque l'algorithme SHAKE est appliqué à toutes les longueurs des liaisons covalentes de la décaalanine. On voit que des déplacements aussi importants que lors d'une dynamique traditionnelle ont lieu le long de cet axe (Fig. II-8a), alors que les longueurs des liaisons N-H ne varient pas au cours de cette simulation. Les déplacements le long de cet axe sont donc pour l'essentiel dûs aux changements de l'orientation dans l'espace des liaisons N-H lors de la dynamique traditionnelle.

Lors d'une simulation de dynamique moléculaire confinée 0-1800, les longueurs des liaisons N-H ne sont pas maintenues constantes. En effet, c'est la différence entre la coordonnée de l'atome H et celle de l'atome N sur l'axe dont la direction est donnée par l'orientation de la liaison N-H à $t=0$ qui est maintenue constante au cours d'une telle simulation (figure II-9). Ainsi, lorsque les liaisons N-H changent d'orientation, leur longueur varie. Comme la constante de force de ces liaisons est très importante (c'est la plus forte des liaisons d'une protéine), ces élongations représentent une forte résistance à toute modification de l'orientation dans l'espace des liaisons N-H. Cette résistance est la cause de la réduction des fluctuations de la position des atomes lors des dynamiques confinées. L'origine ultime de cette réduction réside donc dans le fait que les axes propres sont définis dans un repère cartésien puisqu'ainsi ces axes n'ont aucun sens physique, si ce n'est lorsque la configuration des atomes est très voisine de celle pour laquelle ils ont été calculés. La réduction des fluctuations de la position des atomes lors des dynamiques confinées est donc bien un effet anharmonique. Cependant, cet effet est lié à la résolution des équations du mouvement dans un repère cartésien et à l'utilisation d'axes propres définis dans ce repère.

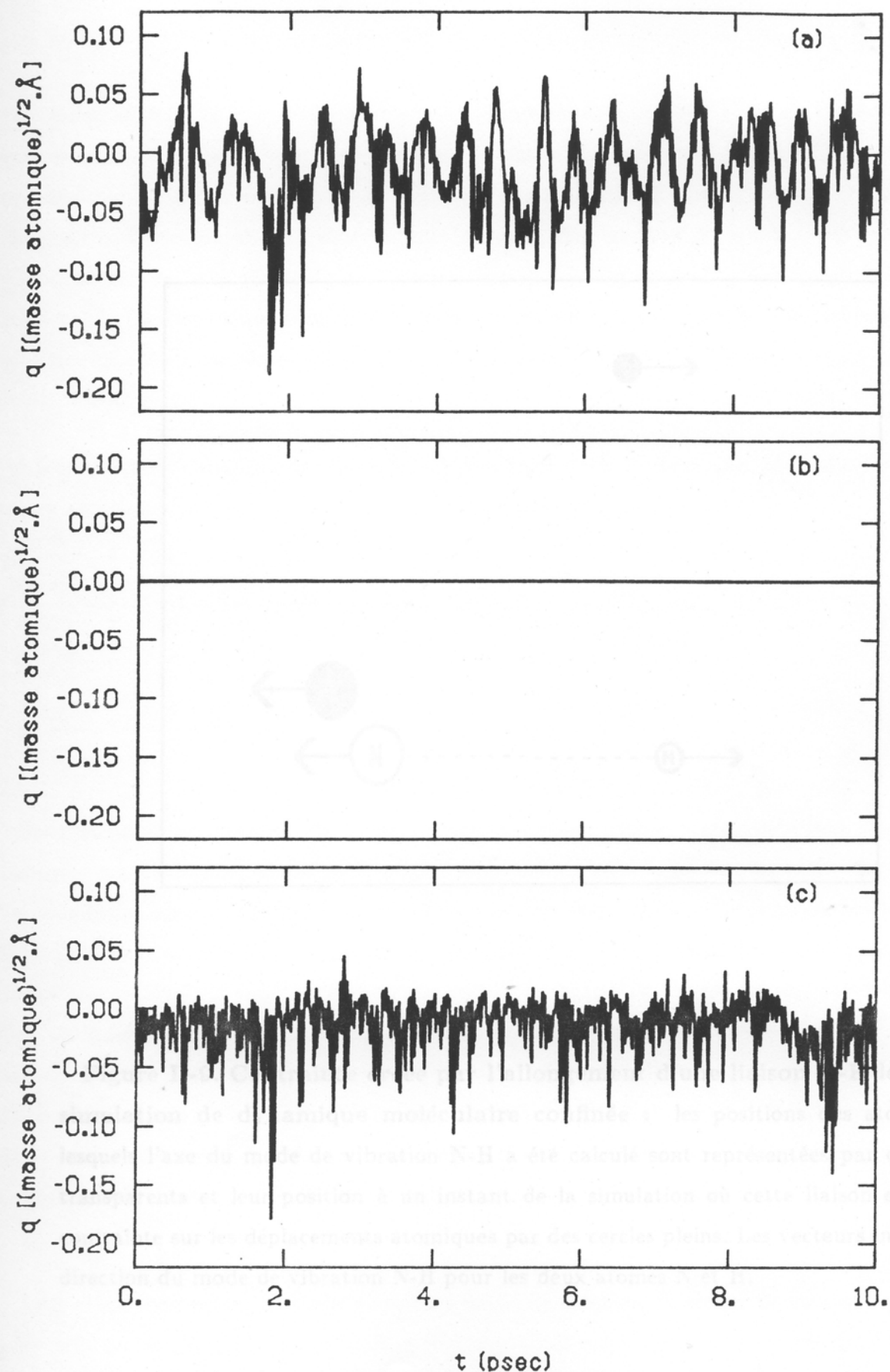


Figure II-8. Projection des déplacements des atomes de la décaalanine à 300°K sur l'axe du mode de vibration de fréquence 3192 cm^{-1} : a) dynamique moléculaire traditionnelle. b) dynamique moléculaire confinée 0-55. c) dynamique moléculaire traditionnelle durant laquelle les longueurs de toutes les liaisons sont maintenues constantes au moyen de l'algorithme SHAKE.

Un autre effet de ce type est mis en évidence lors de chaque simulation de dynamique confinée : le moment angulaire varie au cours du temps. Le tableau II-3 montre les valeurs des composantes du vecteur moment angulaire après 10 psec d'une dynamique confinée 0-55 pour quatre simulations calculées avec des pas d'intégration différents. Ces valeurs dépendent peu de la qualité du pas d'intégration choisi. De plus, elles sont importantes. En effet, à la fin d'une simulation traditionnelle, l'ordre de grandeur de chaque composante du vecteur moment angulaire est d'environ 0,1 [masse atomique].Å²/psec.

Tableau II-3 : Composantes du moment angulaire de la décaalanine, Mx, My, Mz, en [masse atomique].Å²/psec, à la fin d'une dynamique confinée 0-55 de 10 psec.

Pas (psec)	Mx	My	Mz
0,005	-750	610	-140
0,010	-734	617	-159
0,020	-679	645	-224
0,040	-624	757	-300

Pour comprendre l'origine de la non-conservation du moment angulaire lors des dynamiques moléculaires confinées, le plus simple est de raisonner sur un petit exemple. Prenons le cas bien connu d'une molécule triatomique symétrique, telle la molécule d'eau. Cette molécule a trois pseudo-modes de rotation. Lors d'une simulation traditionnelle, lorsqu'une molécule d'eau tourne sur elle même autour d'un axe perpendiculaire à son plan de symétrie, c'est-à-dire dans la direction de l'un de ses trois pseudo-modes de rotation, les atomes sont soumis à une force centrifuge qui excite des modes de vibration de la molécule, notamment le mode d'élongation symétrique des liaisons O-H. L'origine ultime de ce couplage entre le mouvement de rotation globale de la molécule et les vibrations de ses liaisons O-H est la loi de l'inertie : les atomes tendent à se déplacer en ligne droite.

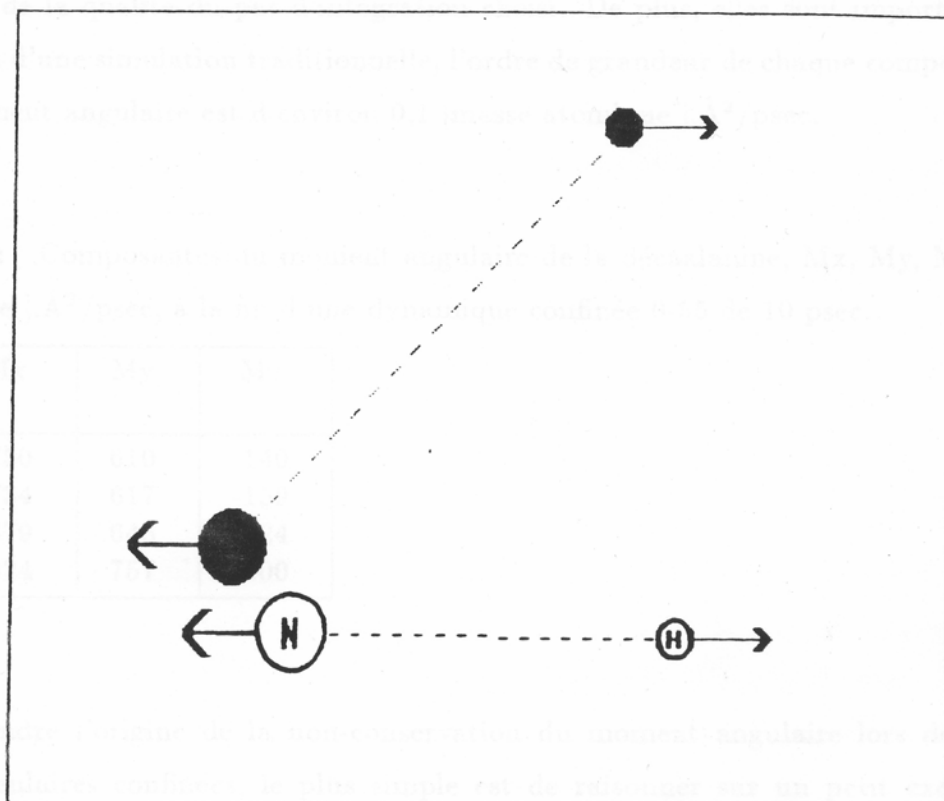
Lors d'une dynamique confinée sur un tel mode de rotation, la valeur du moment angulaire de la molécule, qui n'est pas nulle à $t=0$, ne va pas rester constante comme dans le cas

Un autre effet de ce type est que les vitesses de rotation angulaire de la molécule sont confinées : le moment angulaire total le long de l'axe de rotation est constant. Le tableau II-3 montre les valeurs des composantes du vecteur moment angulaire après 10 ps. D'une dynamique confinée 0-35 ps, quatre composantes ont été calculées avec des pas d'intégration différents. Ces valeurs dépendent peu du pas d'intégration, mais elles sont sensiblement différentes.

En effet, à la fin de la simulation, les vitesses de rotation de chaque composante du vecteur moment angulaire sont d'environ 1.0 unités de moment angulaire.

Tableau II-3 : Composantes du moment angulaire de la molécule, M_x , M_y , M_z , en unités atomiques, $\text{\AA}^2/\text{ps}$, à la fin d'une dynamique confinée à 10 et 35 ps.

Pas (ps)	M_x	M_y	M_z
0.005	1.0	0.9	0.8
0.010	1.0	0.9	0.8
0.020	1.0	0.9	0.8
0.040	1.0	0.9	0.8



Pour compenser l'absence de la non-conservation du moment angulaire lors d'une dynamique confinée, il est nécessaire de choisir un pas d'intégration qui soit un petit multiple.

Prenons le cas bien connu d'une molécule triatomique symétrique, telle la molécule d'eau. Cette molécule a trois périodes de rotation. Lors d'une simulation confinée, lorsqu'une molécule d'eau tourne, elle tourne autour d'un axe perpendiculaire à son plan de symétrie, c'est-à-dire dans la direction de l'un de ses trois axes de rotation.

Figure II-9. Contrainte créée par l'allongement d'une liaison N-H lors d'une simulation de dynamique moléculaire confinée : les positions des atomes pour lesquels l'axe du mode de vibration N-H a été calculé sont représentées par des cercles transparents et leur position à un instant de la simulation où cette liaison exerce une contrainte sur les déplacements atomiques par des cercles pleins. Les vecteurs indiquent la direction du mode de vibration N-H pour les deux atomes N et H.

Lors d'une dynamique confinée, la valeur du moment angulaire de la molécule, qui n'est pas nulle à t=0, ne va pas rester constante. Dans le cas

d'une dynamique moléculaire traditionnelle. Elle va fluctuer. En effet, les déplacements imposés aux atomes par notre méthode sont dans ce cas rectilignes; lorsque les atomes se déplacent le long de l'axe propre de rotation choisi (figure II-10), la longueur des liaisons O-H augmentent, et donc leur énergie aussi. Ainsi se crée une force qui s'oppose au déplacement de la molécule dans le sens initial de sa rotation et la vitesse des atomes diminue avant de s'annuler puis de s'inverser. Durant cette dynamique, au lieu de tourner sur elle même avec une vitesse de rotation constante, la molécule tourne dans un sens puis dans le sens opposé, puis dans le sens initial, oscillant tel un pendule dans un champ de gravitation. Le moment angulaire de la molécule fluctue donc entre 0 et sa valeur initiale, son sens se renversant régulièrement. Cet exemple met clairement en évidence le rôle de contrainte extérieure que joue le choix d'un axe propre défini dans un repère extérieur à la molécule lorsque l'on confine une dynamique.

Il illustre aussi la nature de l'effet anharmonique mis en évidence par la méthode de dynamique moléculaire confinée. En effet, lorsqu'une molécule se déplace le long d'un de ses axes propres, les déplacements de ses atomes sont linéaires. Lorsqu'on applique la théorie des modes normaux à l'étude des propriétés dynamiques d'une molécule, on suppose donc que les mouvements des atomes à une fréquence donnée se font en suivant des lignes droites. Or, lors d'une trajectoire à température ordinaire, ils ne le font pas. On peut se convaincre de cela en considérant l'exemple d'une autre molécule triatomique, CO_2 . Examinons la surface d'énergie potentielle dans laquelle se déplace un des atomes d'oxygène. Dans la direction qui rapproche ou éloigne cet oxygène du carbone, l'énergie potentielle de la molécule augmente rapidement, tandis que dans la direction perpendiculaire à la précédente, elle augmente plus lentement. La surface d'énergie potentielle a donc la forme d'une vallée encaissée dont le fond dessine un arc de cercle (Fig. II-11). En moyenne, lors d'une trajectoire du CO_2 à température ordinaire, les déplacements de l'atome d'oxygène suivent le fond de cette vallée. En d'autres termes, ils sont curvilignes. De fait, cette non-linéarité des déplacements atomiques est à l'origine de l'effet anharmonique mis en évidence par la méthode de dynamique moléculaire confinée. Son importance a été reconnue par d'autres moyens, lors d'études de petites molécules. Elle incite notamment à utiliser des coordonnées curvilignes pour décrire le champ de force des petites molécules

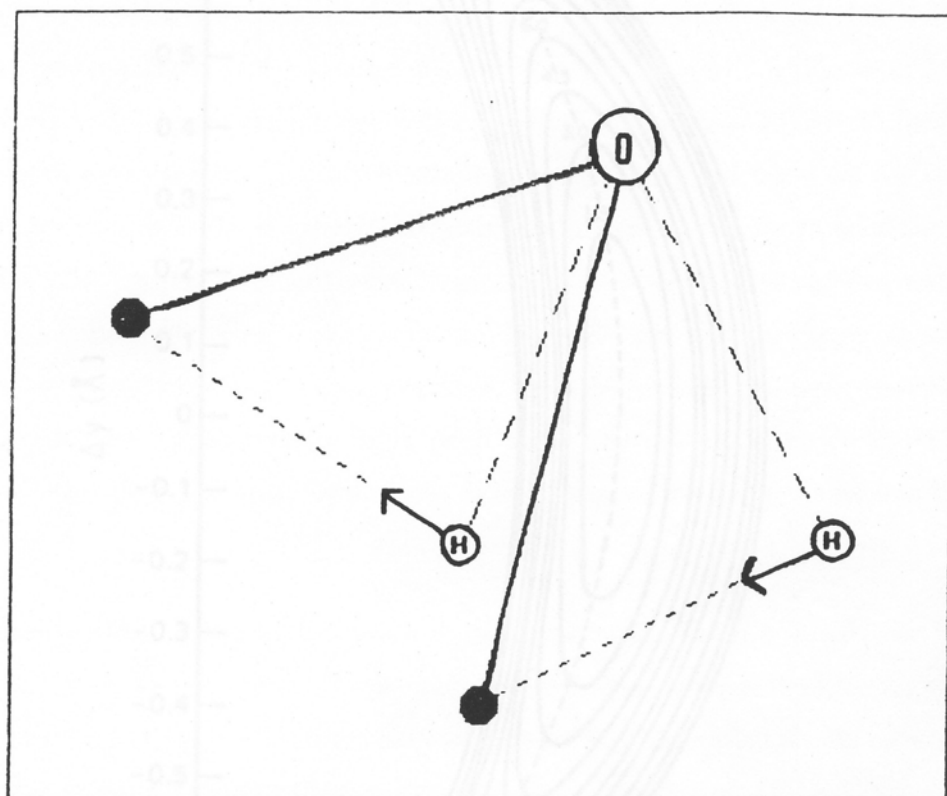


Figure II-10. Origine de la non-conservation du moment angulaire lors d'une simulation de dynamique moléculaire confinée : les positions atomiques pour lesquelles l'axe du pseudo-mode de rotation a été calculé sont représentées par des cercles transparents. Les vecteurs indiquent pour chaque atome la direction de ce pseudo mode de rotation et les cercles pleins montrent la position des atomes à un instant donné d'une simulation de dynamique moléculaire confinée dans le sous-espace défini par l'axe de ce pseudo-mode de rotation.

(I.M.Mills, 1969; A.R.Hoy & col., 1972; I.M.Mills, 1974; S.Califano, 1976; D.Papousek & M.R.Aliev, 1982). Elle justifie les notions d'angle de valence et de liaison covalente, c'est-à-dire leur utilisation dans des fonctions d'énergie potentielle empirique. Elle est aussi à l'origine de ce qu'on appelle la résonance de Fermi, dans le cas de molécules telles H_2O ou CO_2 , c'est-à-dire du couplage entre un mode de vibration des liaisons covalentes et le mode d'ouverture-fermeture de l'angle de valence de la molécule (E.Fermi, 1931; D.M.Dennison, 1940). Pour de grosses molécules, et la décaalanine est pour un chimiste ou un spectroscopiste une très grosse molécule, notre méthode de dynamique moléculaire confinée est de loin la méthode la plus simple pour mettre en évidence et mesurer l'intensité des couplages dûs à cet effet anharmonique, c'est-à-dire à la non-linéarité des déplacements atomiques. Cet effet est intense même à de basses températures, puisque lors des dynamiques confinées 0-200 et 0-800 à 100°K les fluctuations de la position des atomes C_α sont de 23 % moins importantes que celles de ces atomes lors d'une dynamique moléculaire traditionnelle.

Pour tenir compte, dans notre méthode de dynamique moléculaire confinée, de la non-linéarité des déplacements atomiques, une première idée vient à l'esprit : utiliser un système de coordonnées curvilignes pour décrire les axes propres d'une molécule. Un tel ensemble de coordonnées existe et nous l'avons utilisé pour appliquer la méthode GF de Wilson (I-C2) ou la méthode de projection de Williams (I-C3). C'est l'ensemble des coordonnées internes. On pourrait donc généraliser la méthode de dynamique moléculaire confinée en utilisant des axes propres définis en termes de coordonnées internes, c'est-à-dire de variation de longueur de liaisons covalentes, d'écartement d'angles de valence et d'angles dièdres. Une telle généralisation n'est cependant pas simple à mettre en oeuvre. En effet, lorsque la structure de la molécule étudiée s'éloigne un tant soit peu de la structure pour laquelle les matrices B , T , G , et A_I ont été calculées, les directions des axes propres, telles qu'elles sont décrites dans A_I , ne sont plus indépendantes. En d'autres termes, A_I n'est que localement orthogonale. D'autre part, si la matrice A_I est régulièrement "orthogonalisée", les équations du mouvement dans le repère propre deviennent plus complexes à résoudre puisque la matrice A_I varie au cours du temps (cf partie I-E). La généralisation de la méthode de dynamique moléculaire confinée ne peut donc se faire par cette voie qu'au prix d'une complexification très importante et il est fort probable qu'une telle complexification

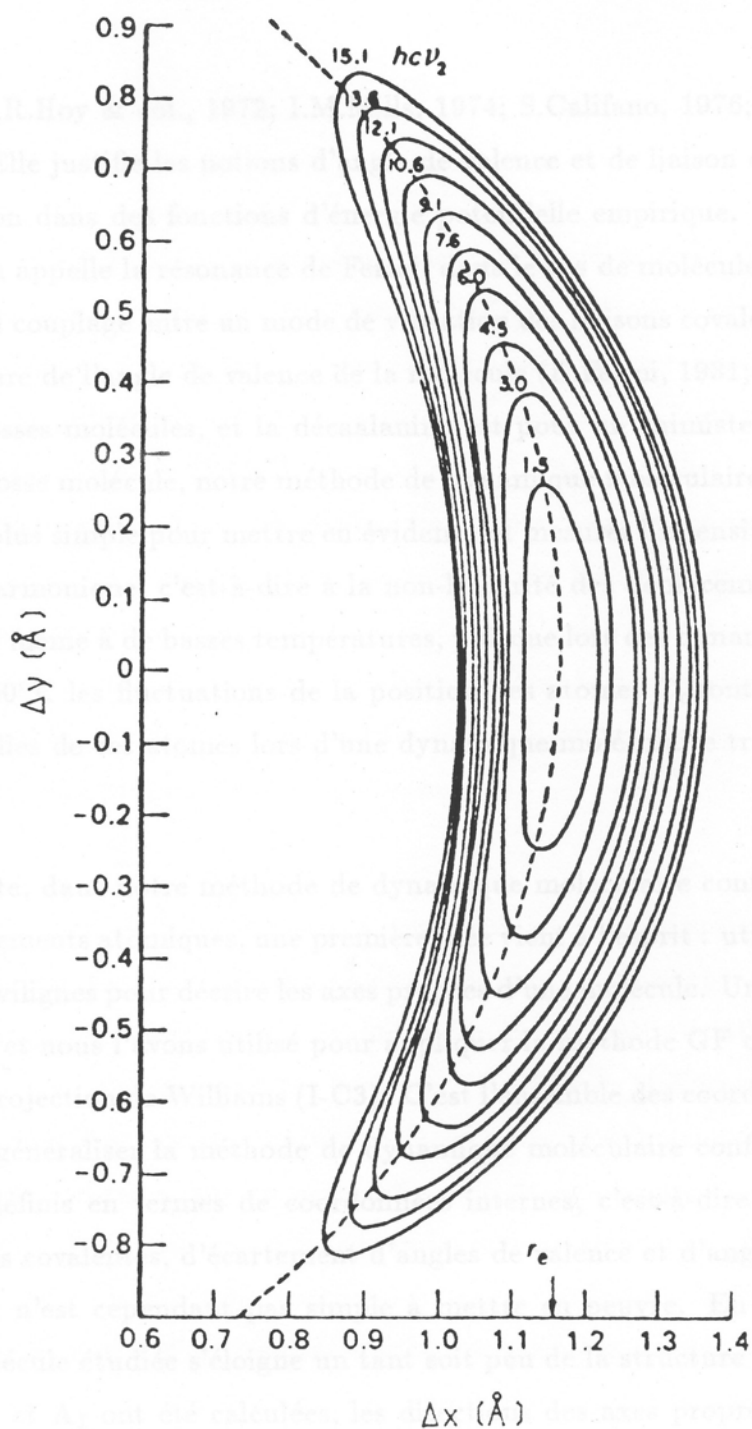


Figure II-11. Surface d'énergie potentielle vue par un des atomes d'oxygène de la molécule CO_2 . L'axe x est l'axe de symétrie de la molécule; r_e est la distance moyenne entre l'oxygène et le carbone (M.A.Pariseau & col., 1965; I.M.Mills, 1974).

rendrait la méthode trop lourde à mettre en oeuvre.

On peut songer à d'autres voies de développement de notre méthode comme, par exemple, faire alterner des périodes de dynamique confinée, calculées avec un grand pas d'intégration, avec des périodes de dynamique traditionnelle durant lesquelles une partie seulement de l'énergie potentielle est prise en compte, ceci par analogie avec la méthode à "multiple pas d'intégration" (W.B.Streett & col., 1978). L'idée sous-jacente à ce procédé est de laisser régulièrement se relacher les tensions qui s'accumulent lorsque les déplacements des atomes sont confinés dans un sous-espace donné. Le résultat qui justifie une telle approche est celui mis en évidence par notre méthode de dynamique moléculaire projetée, à savoir que lors d'une dynamique traditionnelle de la décaalanine à température ordinaire, les atomes se déplacent dans un sous-espace convenablement décrit par une partie seulement des axes propres de la décaalanine.

Cependant, cette voie de développement ultérieur de la méthode de dynamique moléculaire confinée est suggérée par des résultats obtenus en étudiant des dynamiques durant lesquelles aucun changement de conformation ne se produit. Or, une protéine change de conformation environ toutes les 0,15 psec (R.Elber & M.Karplus, 1987). Avant toute chose, il nous faut donc étudier l'incidence de cet autre effet fortement anharmonique, le changement de conformation, sur les perspectives de développement de notre méthode. Pour cela, nous avons choisi une approche opposée, mais complémentaire, à celle utilisée jusqu'à ce point de notre travail. Pour concevoir notre méthode de dynamique moléculaire confinée, nous sommes partis d'un résultat mathématique, à la fois rigoureux et approximatif. Rigoureux, parce que les équations du mouvement des atomes sont résolues de manière exacte. Approximatif, parce que les solutions de ces équations ne sont valables que lorsque les atomes se déplacent très peu, lors d'une trajectoire à basse température, par exemple. Pour développer plus avant notre méthode, nous souhaitons nous inspirer d'un cas plus réaliste. Pour ce faire, nous allons, dans la dernière partie de notre travail, étudier des trajectoires de protéines à température ordinaire. Dans ces trajectoires, des mouvements amples et périodiques sont plus particulièrement étudiés, c'est-à-dire des

mouvements dont les caractéristiques sont proches de celles des modes de vibration de basse fréquence. La question à laquelle nous cherchons à répondre lors de ces études est la suivante : y-a-t-il une relation nette entre les mouvements amples et périodiques des atomes d'une protéine à très basse température et ceux à température ordinaire ? Si une telle relation existe, elle nous servira de guide pour définir les voies de développement ultérieur de la méthode de dynamique moléculaire confinée.

Des déplacements de boucles, d'hélices α , voire de domaines entiers, ont été mis en évidence lors de l'expression de l'activité de nombreuses protéines (J.J. Holbrook & col., 1979; H. Fehlbammier & col., 1977; C.M. Anderson & col., 1979; E.W. Dijkstra & col., 1982). Ces mouvements relatifs d'amplitude importante ont surtout été détectés en comparant les structures de protéines cristallisées en absence ou en présence de leur substrat. Certaines simulations de dynamique moléculaire ont permis de mettre en évidence des mouvements amples impliquant des parties structurées entières d'une protéine. Le cas le plus net concerne le mouvement périodique relatif des hélices αB et αC du fragment C-terminal de la protéine ribosomale L7/L12 d'*E. Coli* (CTF) pour lequel la fréquence (environ 5 cm^{-1}) et l'amplitude (environ 15°) ont été déterminées lors de l'analyse d'une trajectoire de 150 psec du monomère dans le vide (J. Åqvist & col., 1985). Ce mode de mouvement perdure dans des conditions environnementales très diverses. En effet, un mouvement périodique de fréquence et d'amplitude comparables a été retrouvé lors de l'analyse d'une simulation d'un dimère de ce fragment (J. Åqvist, 1987), puis lors de celle du monomère en présence d'eau (O. Tapia, 1989).

Au sein du ribosome, ou en solution, la protéine L7/L12 est sous forme de dimère (W. Moller & col., 1972; R. Osterberg & col., 1976). Deux dimères de cette protéine sont liés à la grande sous-unité (50 S) du ribosome via la protéine L10 (A.T. Gudkov & col., 1980). La protéine L7/L12 est essentielle à la synthèse polypeptidique dans la bactérie (W. Moller & col., 1974; I. Petterson et C.G. Kurland, 1980). Elle est constituée de 120 acides aminés. Sa partie N-terminale (résidus 1 à 52) est impliquée dans la formation du dimère (W. Moller & col., 1972) et dans la fixation sur le ribosome (A.J. Van Agthoven & col., 1975; V.E. Kotliansky & col., 1978; J.A. Maassen et W. Moller, 1981). La partie

III. Etude des mouvements de grande amplitude d'une petite protéine, le fragment C-terminal de la protéine ribosomale L7/L12

Des déplacements de boucles, d'hélices α , voire de domaines entiers, ont été mis en évidence lors de l'expression de l'activité de nombreuses protéines (J.J.Holbrook & col., 1975; H.Fehlhammer & col., 1977; C.M.Anderson & col., 1979; B.W.Dijkstra & col., 1982). Ces mouvements relatifs d'amplitude importante ont surtout été détectés en comparant les structures de protéines cristallisées en absence ou en présence de leur substrat. Certaines simulations de dynamique moléculaire ont permis de mettre en évidence des mouvements amples impliquant des parties structurées entières d'une protéine. Le cas le plus net concerne le mouvement périodique relatif des hélices αB et αC du fragment C-terminal de la protéine ribosomale L7/L12 d'E.Coli (CTF) pour lequel la fréquence (environ 5 cm^{-1}) et l'amplitude (environ 15°) ont été déterminées lors de l'analyse d'une trajectoire de 150 psec du monomère dans le vide (J.Åqvist & col., 1985). Ce mode de mouvement perdure dans des conditions environnementales très diverses. En effet, un mouvement périodique de fréquence et d'amplitude comparables a été retrouvé lors de l'analyse d'une simulation d'un dimère de ce fragment (J.Åqvist, 1987), puis lors de celle du monomère en présence d'eau (O.Tapia, 1990).

Au sein du ribosome, ou en solution, la protéine L7/L12 est sous forme de dimère (W.Moller & col., 1972; R.Osterberg & col., 1976). Deux dimères de cette protéine sont liés à la grande sous-unité (50 S) du ribosome via la protéine L10 (A.T.Gudkov & col., 1980). La protéine L7/L12 est essentielle à la synthèse polypeptidique dans la bactérie (W.Moller & col., 1974; I.Petterson et C.G.Kurland, 1980). Elle est constituée de 120 acides aminés. Sa partie N-terminale (résidus 1 à 52) est impliquée dans la formation du dimère (W.Moller & col., 1972) et dans la fixation sur le ribosome (A.J.Van Aghoven & col., 1975; V.E.Koteliarsky & col., 1978; J.A.Maasen et W.Moller, 1981). La partie

C-terminale (résidus 53 à 120) semble être la partie fonctionnelle puisque sa présence est nécessaire à l'activité GTP-ase des facteurs d'élongation EF-Tu et EF-G (D.M.Marquis et S.R.Fahnestock, 1980). La protéine L7/L12 a une assez large mobilité sur le ribosome et ne parait pas impliquée dans l'association des sous-unités ribosomales (W.Moller & col., 1983). Sa flexibilité est perdue lorsqu'un complexe formé du facteur d'élongation EF-G et d'un analogue du GTP, le GDPCP, est fixé sur le ribosome (A.T.Gudkov et G.M.Gongadze, 1984). Ainsi, un changement conformationnel pourrait être à la base du fonctionnement de la protéine L7/L12 durant le processus de traduction (W.Moller & col., 1983). Ce changement devrait affecter la partie fonctionnelle de la protéine, c'est-à-dire son domaine C-terminal (CTF).

Comme l'ont montré les expériences de dénaturation par le pH, la température et le chlorure de guanidine, le CTF est un petit fragment compact très stable (C.A.Luer et K.P.Wong, 1980). Sa structure a été déterminée avec une résolution de 1,7 Å (M.Leijonmarck & col., 1980; M.Leijonmarck et A.Liljas, 1987). Le taux de résidus appartenant à des hélices α est compris entre 46% et 59% selon que l'on prend pour critère les valeurs des angles Φ et Ψ ou le réseau des liaisons hydrogènes formées (J.Åqvist & col., 1985). Il y a trois hélices placées côte à côte, l'hélice centrale (αA) étant antiparallèle aux deux autres, et trois feuillets β antiparallèles (Fig. III-1). Cette abondance de structures secondaires facilite la mise en évidence de mouvements atomiques collectifs puisqu'on peut, dans un premier temps, se contenter d'étudier leurs mouvements relatifs.

Il a souvent été suggéré que des mouvements atomiques collectifs d'origine thermique pourraient avoir un rôle fonctionnel majeur (F.R.N.Gurd & T.M.Rothgeb, 1979; R.J.P.Williams, 1979; M.Karplus & J.A.McCammon, 1981; 1983). Or, contrairement aux petites protéines qui ont servi de modèles lors du développement des méthodes de dynamique moléculaire, telles le BPTI et la myoglobine, le CTF participe à des activités catalytiques. Ce n'est donc peut être pas un hasard si un mouvement périodique de grande amplitude et de basse fréquence a été mis en évidence lors de chacune des simulations du CTF. Quoiqu'il en soit, le CTF est un bon modèle pour approfondir la question de la

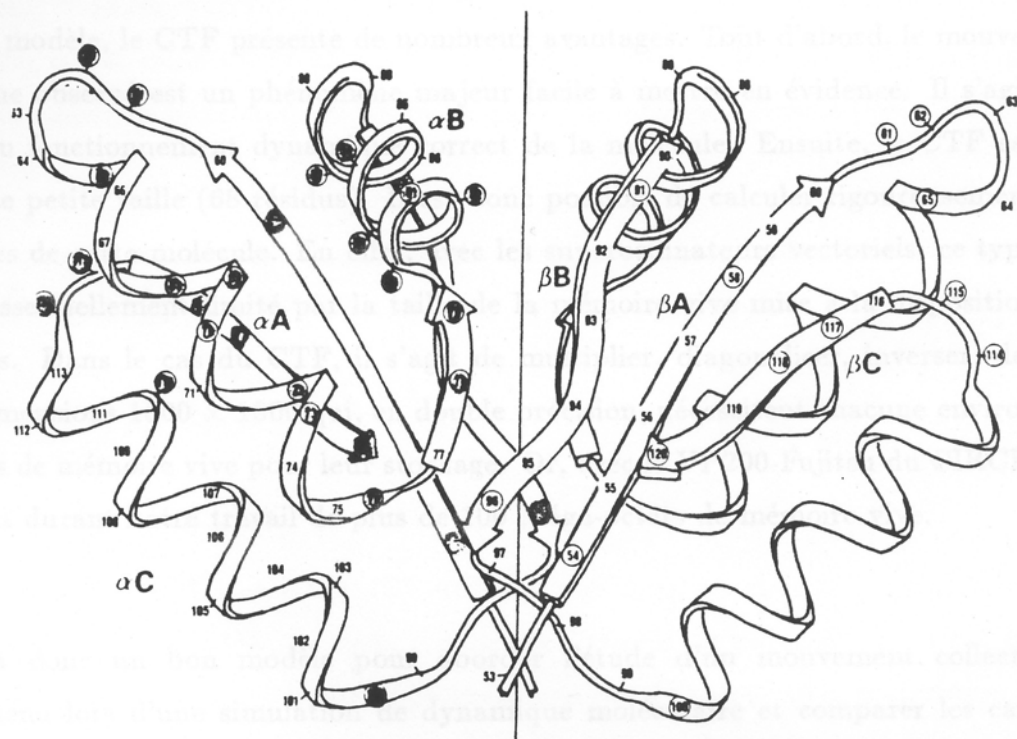


Figure III-1. Structure cristallographique du dimère du fragment C-terminal (résidus 53 à 120) de la protéine ribosomale L7/L12 d'E.Coli (A.Liljas & M.Leijonmarck, 1983; J.Aqvist & col., 1985; reproduction faite avec l'autorisation du *Journal of Molecular Biology*).

relation entre les mouvements d'origine thermique et l'activité fonctionnelle des protéines.

En tant que modèle, le CTF présente de nombreux avantages. Tout d'abord, le mouvement périodique observé est un phénomène majeur facile à mettre en évidence. Il s'agit d'un témoin du fonctionnement dynamique correct de la molécule. Ensuite, le CTF est une protéine de petite taille (68 résidus). Il est donc possible de calculer rigoureusement les axes propres de cette molécule. En effet, avec les superordinateurs vectoriels, ce type de calcul est essentiellement limité par la taille de la mémoire vive mise à la disposition des utilisateurs. Dans le cas du CTF, il s'agit de multiplier, diagonaliser, inverser, des matrices de dimensions 1800×1800 qui, en double précision, nécessitent chacune environ 32 Méga-octets de mémoire vive pour leur stockage. Or, avec le VP200-Fujitsu du CIRCE, nous disposions durant notre travail de plus de 100 Méga-octets de mémoire vive.

Le CTF est donc un bon modèle pour aborder l'étude d'un mouvement collectif périodique obtenu lors d'une simulation de dynamique moléculaire et comparer les caractéristiques de ce mouvement à celles des mouvements périodiques prédits par la théorie des modes normaux.

L'existence d'un phénomène témoin va permettre, en outre, de vérifier que les résultats obtenus par les méthodes de dynamique moléculaire sont reproductibles. En effet, toutes les simulations de dynamique moléculaire du CTF faites à ce jour l'ont été avec le programme GROMOS (W.F. Van Gunsteren & col., 1983; J. Åqvist & col., 1985), c'est-à-dire avec des options méthodologiques bien particulières. Or, nous travaillons avec un autre programme, CHARMM, réalisé de manière indépendante, avec une philosophie comparable cependant. Retrouver ce phénomène avec cet autre programme permettrait de lever une des réserves que l'on peut faire sur la validité de résultats obtenus par les méthodes de dynamique moléculaire. En effet, à notre connaissance, les comparaisons de trajectoires calculées avec deux programmes différents n'ont été faites à ce jour que sur la base de critères très globaux : fluctuations moyennes de la position des atomes, de la valeur des angles dièdres,

etc... On est donc en droit de s'interroger sur la généralité d'un résultat obtenu avec un seul programme, c'est-à-dire, notamment, avec un seul ensemble de paramètres. Nous avons donc voulu, O.Tapia et moi-même, vérifier, dans un premier temps, que le mouvement relatif des hélices αB et αC du CTF découvert par Åqvist est peu sensible aux détails de la méthodologie employée pour calculer les trajectoires des atomes de cette protéine.

A. Comparaison des surfaces d'énergie potentielle calculées par les programmes CHARMM et GROMOS.

1) topologie.

Les programmes CHARMM et GROMOS utilisent une fonction énergie potentielle de même forme (voir chapitre I-B) mais deux ensembles distincts de paramètres. Ainsi, les surfaces de potentiel explorées par les deux programmes lors du calcul de la trajectoire du CTF sont différentes, et ce, d'autant plus que la description de la chaîne latérale de certains résidus n'est pas la même dans les deux programmes. Par exemple, pour GROMOS, le CTF a 263 angles dièdres impropres et 336 angles dièdres (Åqvist & col., 1985), tandis que pour CHARMM il en a respectivement 239 et 360. L'énergie potentielle de 24 des 599 angles dièdres du CTF est donc calculée de manière très différente selon que l'on utilise l'un ou l'autre des deux programmes.

2) Charges partielles.

Le CTF possède de nombreux résidus chargés, 10 lysines, 1 arginine, 9 glutamates et 4 aspartates. Or, la valeur absolue de l'énergie potentielle d'une protéine dépend beaucoup des options méthodologiques choisies pour calculer le terme d'interaction électrostatique. Nous avons donc choisi de consacrer une étude séparée à l'influence des conditions du calcul du terme électrostatique sur le déroulement d'une trajectoire.

Lors d'une première simulation, nous cherchons à reproduire les résultats obtenus par Åqvist. Le terme électrostatique de la fonction énergie potentielle du CTF est donc calculé dans des conditions aussi voisines que possibles des siennes. Ainsi, les charges partielles des groupes chargés NH_3^+ et COO^- ont été modifiées de façon à neutraliser la charge globale de ces groupes, tout en maintenant leur capacité à participer à des liaisons hydrogènes; au total, les charges de 129 des 596 atomes du CTF ont été modifiées (J.Åqvist & col., 1985). Ces modifications miment les effets d'écran dûs à la polarisation du milieu environnant et à la présence de contre ions dans ce milieu (R.Constanciel et O.Tapia, 1978; O.Tapia, 1980). De plus, le comportement artefactuel des chaînes latérales des résidus chargés, qui se replient vers le centre de la protéine lors d'une simulation réalisée en l'absence de molécules de solvant, tend à disparaître. Pour plus de simplicité, on a pris pour valeur de ces charges atomiques partielles modifiées celles utilisées par Åqvist.

3) Milieu environnant.

Durant le calcul des trajectoires du CTF par le programme GROMOS, la température était maintenue constante, à environ 277°K. Cette condition simule la présence d'un bain thermique extérieur faiblement couplé au comportement dynamique de la molécule (W.F.Van Gunsteren & col., 1983; H.J.C.Berendsen & col., 1984). Un tel bain thermique peut jouer un rôle dans la genèse ou l'entretien d'un mouvement périodique. Or, durant le calcul de notre trajectoire par le programme CHARMM, l'énergie totale du CTF est constante, aux erreurs dues à l'algorithme utilisé et à celles dues à l'ordinateur près. En comparant les trajectoires calculées par CHARMM et GROMOS, nous allons donc pouvoir mesurer l'influence de la présence d'un bain thermique extérieur sur le comportement dynamique intrinsèque d'une protéine.

4) Calcul de l'énergie électrostatique

Lors des précédentes simulations, faites avec le programme GROMOS, les interactions

électrostatiques et de Van der Waals entre atomes séparés par une distance supérieure à une distance donnée (8 Å), dite de coupure, n'étaient pas calculées, ceci afin de diminuer le coût en CPU du calcul de l'énergie potentielle du CTF. Il n'est pas recommandé d'utiliser cette option méthodologique lorsqu'on désire calculer une trajectoire à énergie totale constante. En effet, les fluctuations de l'énergie totale, dues au fait que des atomes traversent la parois de la sphère définie autour d'un atome donné par la distance de coupure, peuvent s'amplifier et perturber le comportement dynamique simulé. Pour calculer une trajectoire à énergie totale constante dans des conditions mimant les conditions de calcul des simulations précédentes, réalisées à température constante, les interactions entre atomes non liés sont calculées ainsi : la distance de coupure est de 9,0 Å; pour les atomes distants de 7,0 à 9,0 Å, la valeur de l'énergie potentielle d'interaction de type Van der Waals est ramenée artificiellement à 0 (coupure de type SWITCH; voir B.Brooks & col., 1983) tandis que celle de l'interaction de type électrostatique est rendue nulle lorsque la distance est de 9,0 Å (coupure de type SHIFT). Ces formules de coupure des interactions entre atomes non liés, de type SWITCH et SHIFT, assurent la continuité de la fonction énergie potentielle d'interaction pour des distances voisines de la distance de coupure. Ainsi, les fluctuations de l'énergie totale de la protéine, dues à la coupure des effets du potentiel à partir d'une distance donnée, sont réduites. Comme lors de la simulation d'Åqvist, le critère de coupure est appliqué à la distance qui sépare les centres de géométrie des groupes neutres d'atomes. Enfin, comme précédemment, la constante diélectrique, ϵ_r , est posée égale à 1.

B. Minimisation

L'énergie potentielle de la structure cristallographique du CTF (facteur R de 17%) est soigneusement minimisée au moyen de l'algorithme POWELL. Après environ 4000 étapes, l'algorithme converge. Afin de calculer avec une bonne précision les axes propres de la structure obtenue, le critère de convergence adopté est très rigoureux. En effet, la minimisation est poursuivie jusqu'à ce que la valeur du gradient moyen soit inférieure ou égale à 0,0001 Kcal/Mole.Å. Durant la minimisation, l'énergie potentielle du CTF passe de -439,81 Kcal/Mole à -2380,93 Kcal/Mole; 85 % de ce gain énergétique provient de réarrangements

locaux de la configuration atomique (longueurs de liaison, angles de valence, angles dièdres impropres, etc...). De tels réarrangements ne sont pas surprenants. En effet, l'analyse de la diffraction des rayons X par une molécule cristallisée donne l'image de la structure moyenne de cette molécule et non l'image d'une structure dont l'énergie potentielle est minimale. En d'autres termes, la structure obtenue par ce type de méthode expérimentale est une superposition de toutes les configurations atomiques que la molécule peut adopter au sein d'un cristal.

La distance moyenne entre notre structure et la structure cristallographique est de 0.71 Å si l'on ne tient compte que des atomes C α , et de 1.25 Å si l'on tient compte de tous les atomes. Les angles que font les axes des hélices du CTF entre eux changent de 2 à 4° durant la minimisation. Par ailleurs, le système des liaisons hydrogènes de ces hélices n'est pas modifié. La structure dont les propriétés dynamiques sont étudiées ci-dessous est donc très peu différente de la structure cristallographique.

C. Calcul des fréquences et des axes propres.

La matrice des constantes de force, **F**, est calculée en double précision par la méthode des différences finies avec un pas de 10^{-5} Å. On a vérifié, en comparant la valeur du gradient résiduel moyen calculé de manière analytique à celle obtenue par différences finies, que 10^{-5} Å est une valeur de pas adéquate lorsque la méthode des différences finies est appliquée à une protéine. L'ordre de grandeur de la différence entre ces deux valeurs est de 10^{-9} Kcal/Mole.Å.

Dans un premier temps, afin d'évaluer si la méthode de projection de Williams, nécessaire en toute rigueur, l'est aussi en pratique, la matrice **F** n'est pas projetée. Elle est directement diagonalisée au moyen de la routine vectorisée DSEIG1 (méthode QL) de la librairie mathématique du VP200. Cette opération prend moins de quatre minutes de CPU. Les 6 pseudo-fréquences propres et les 4 plus basses fréquences propres obtenues ont pour

valeurs : $i0.0091$, $i0.0014$, $i0.0011$, 0.0017 , 0.0083 , 0.0218 , 4.77 , 6.11 , 7.10 , et 9.65 cm^{-1} , respectivement.

On a ensuite projeté **F** après avoir calculé **B** et **T** pour les 3N-6 coordonnées internes définies dans CHARMM. A la suite de cette opération, les 6 pseudo-fréquences propres ont pour valeur : 0.00015 , 0.00017 , 0.00020 , 0.00022 , 0.00024 , 0.00041 cm^{-1} , respectivement. Dans ce cas, leur non-nullité peut être attribuée aux erreurs de calcul de l'ordinateur. L'influence de la projection de Williams sur la valeur des fréquences propres du CTF est négligeable, de l'ordre de 10^{-7} cm^{-1} , ce qui confirme, à posteriori, la qualité de la minimisation décrite ci-dessus. Les directions des axes propres calculés sont aussi peu significativement modifiées par l'opération de projection. En absence de projection, la contamination rotationnelle moyenne de ces axes est de $6.10^{-6} \%$. Après projection, elle est de $3.10^{-11} \%$, c'est-à-dire de l'ordre de grandeur des erreurs de calcul de l'ordinateur. Comme dans le cas de la décaalanine, la méthode de Williams ne s'avère donc pas nécessaire pour calculer avec une bonne précision les modes de vibration d'une molécule dont la fonction énergie potentielle a été convenablement minimisée.

D. Première simulation

1) Conditions générales de l'étude.

Notre trajectoire du CTF est calculée avec l'algorithme de Verlet, au lieu du "leap-frog" (R.W.Hockney & J.W.Eastwood, 1981) utilisé par Åqvist pour calculer sa trajectoire de 150 psec. Ces deux algorithmes sont strictement équivalents (H.J.C.Berendsen & W.F.Van Gunsteren, 1984).

Par ailleurs, la liste des atomes non liés en interaction est renouvelée toutes les 5 fsec. Elle donne l'ensemble des paires d'atomes qui sont situées à moins de $10,5 \text{ Å}$ les uns des

autres. L'algorithme SHAKE est appliqué, à chaque étape de la dynamique, à toutes les liaisons covalentes; leur longueur sont maintenues constantes à 10^{-3} Å près. Après une période initiale de 1 psec, durant laquelle le pas d'intégration est de 0.5 fsec, une trajectoire de 65 psec est calculée avec un pas de 1 fsec.

2) Conditions initiales.

La structure à $t=0$ est la structure à partir de laquelle ont été calculées les fréquences et les axes propres du CTF. Afin d'obtenir une trajectoire dont la température moyenne est proche de la température du bain thermique des trajectoires calculées avec GROMOS, 277°K, les vitesses des atomes à $t=0$ ont été calculées de façon à ce que la température initiale du système soit de 540°K. Comme lors du calcul des trajectoires de la décaalanine, ces vitesses sont affectées aux atomes de façon à ne favoriser initialement aucun mode de vibration particulier. Il n'y a pas de période de thermalisation lors du calcul de cette trajectoire.

3) Analyse de la trajectoire obtenue.

La figure III-2a montre l'évolution en fonction du temps de la distance moyenne séparant la structure du CTF de sa structure cristallographique, d'une part, lorsque tous les atomes sont pris en compte, et d'autre part lorsque seuls les C_α le sont. Cette figure est à comparer à celle obtenue par Åqvist (Fig. III-2b). Ces deux figures sont très voisines. Durant les 20 premières picosecondes, la distance évolue vers une valeur limite autour de laquelle elle fluctue par la suite. A l'instar d'Åqvist, on peut donc considérer qu'après 20 psec de dynamique, la période d'équilibration du système est achevée; les analyses porteront donc sur les 45 psec ultérieures. Les distances moyennes durant cette dernière période sont un peu moins grandes que celles trouvées par Åqvist, alors que l'allure des fluctuations est très similaire.

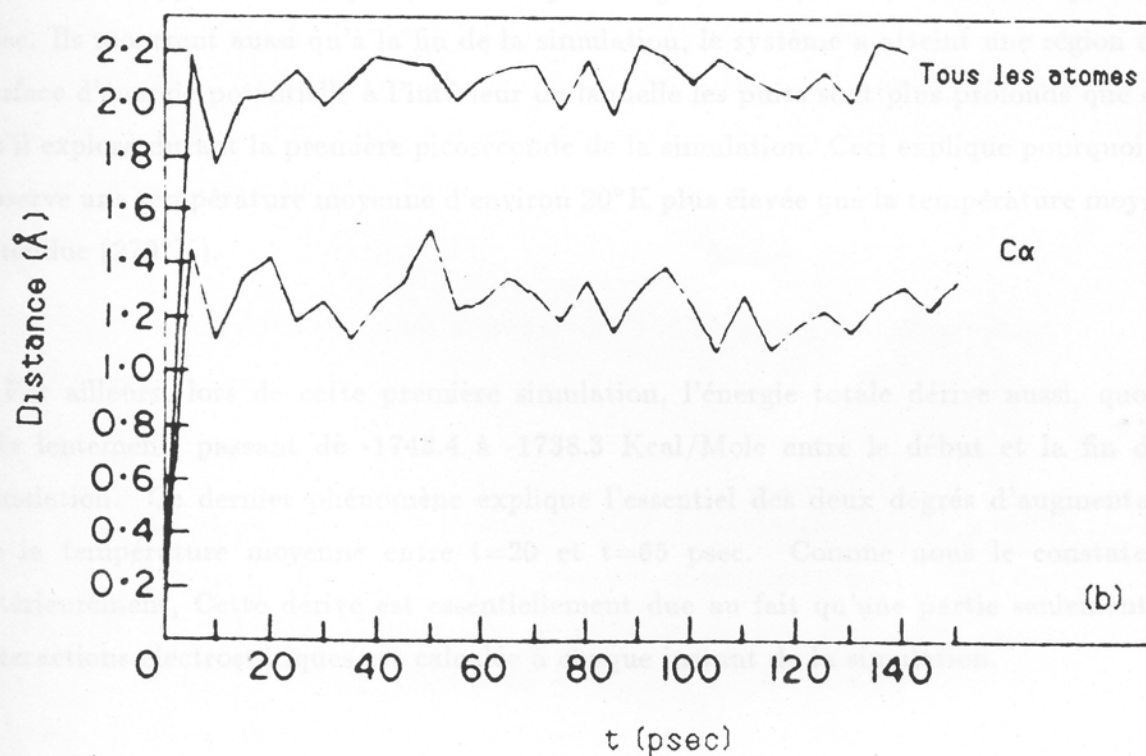
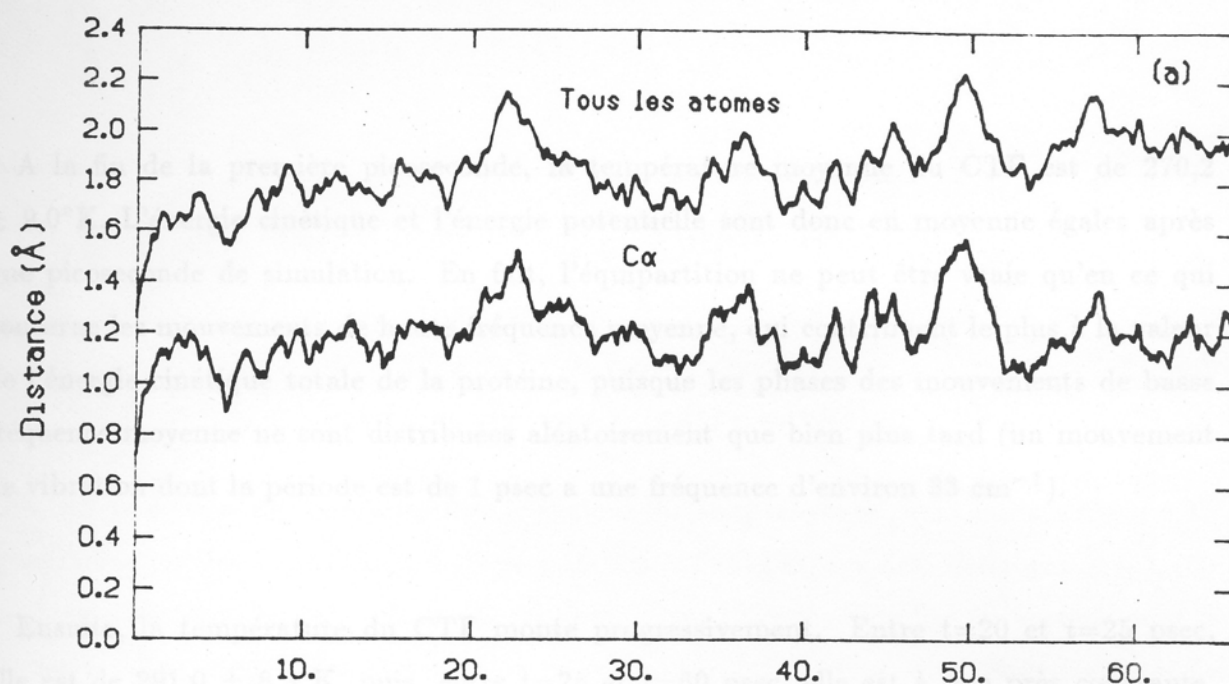


Figure III-2. Evolution en fonction du temps de la distance moyenne séparant la structure instantanée du CTF de sa structure cristallographique. Dans la courbe supérieure, tous les atomes sont pris en compte lors du calcul de cette distance, alors que dans la courbe inférieure seuls les atomes C_{α} le sont. (a) trajectoire calculée avec CHARMM. (b) trajectoire calculée avec GROMOS (J.Aqvist & col., 1985; reproduction faite avec l'autorisation du *Journal of Molecular Biology*).

A la fin de la première picoseconde, la température moyenne du CTF est de $270,2 \pm 9,0^\circ\text{K}$. L'énergie cinétique et l'énergie potentielle sont donc en moyenne égales après une picoseconde de simulation. En fait, l'équipartition ne peut être vraie qu'en ce qui concerne les mouvements de haute fréquence moyenne, qui contribuent le plus à la valeur de l'énergie cinétique totale de la protéine, puisque les phases des mouvements de basse fréquence moyenne ne sont distribuées aléatoirement que bien plus tard (un mouvement de vibration dont la période est de 1 psec a une fréquence d'environ 33 cm^{-1}).

Ensuite, la température du CTF monte progressivement. Entre $t=20$ et $t=25$ psec, elle est de $291,0 \pm 8,3^\circ\text{K}$, puis, entre $t=25$ et $t=60$ psec, elle est à peu près constante, atteignant, entre $t=60$ et $t=65$ psec, une valeur moyenne de $293,9 \pm 8,7^\circ\text{K}$. Ces résultats confirment l'hypothèse d'Åqvist, à savoir qu'un régime stationnaire est atteint après $t=20$ psec. Ils montrent aussi qu'à la fin de la simulation, le système a atteint une région de la surface d'énergie potentielle à l'intérieur de laquelle les puits sont plus profonds que celui qu'il explore durant la première picoseconde de la simulation. Ceci explique pourquoi l'on observe une température moyenne d'environ 20°K plus élevée que la température moyenne attendue (270°K).

a) Comparaison de la structure moyenne et de la structure cristallographique.

Par ailleurs, lors de cette première simulation, l'énergie totale dérive aussi, quoique très lentement, passant de -1743.4 à -1738.3 Kcal/Mole entre le début et la fin de la simulation. Ce dernier phénomène explique l'essentiel des deux degrés d'augmentation de la température moyenne entre $t=20$ et $t=65$ psec. Comme nous le constaterons ultérieurement, Cette dérive est essentiellement due au fait qu'une partie seulement des interactions électrostatiques est calculée à chaque instant de la simulation.

Une représentation bidimensionnelle de la trajectoire permet de visualiser l'évolution globale du CTF. Deux coordonnées générales, quelque peu arbitraires, sont sélectionnées dans cette perspective. La première est la distance moyenne entre la structure du CTF à un instant t et sa structure cristallographique, calculée en ne tenant compte que des seuls

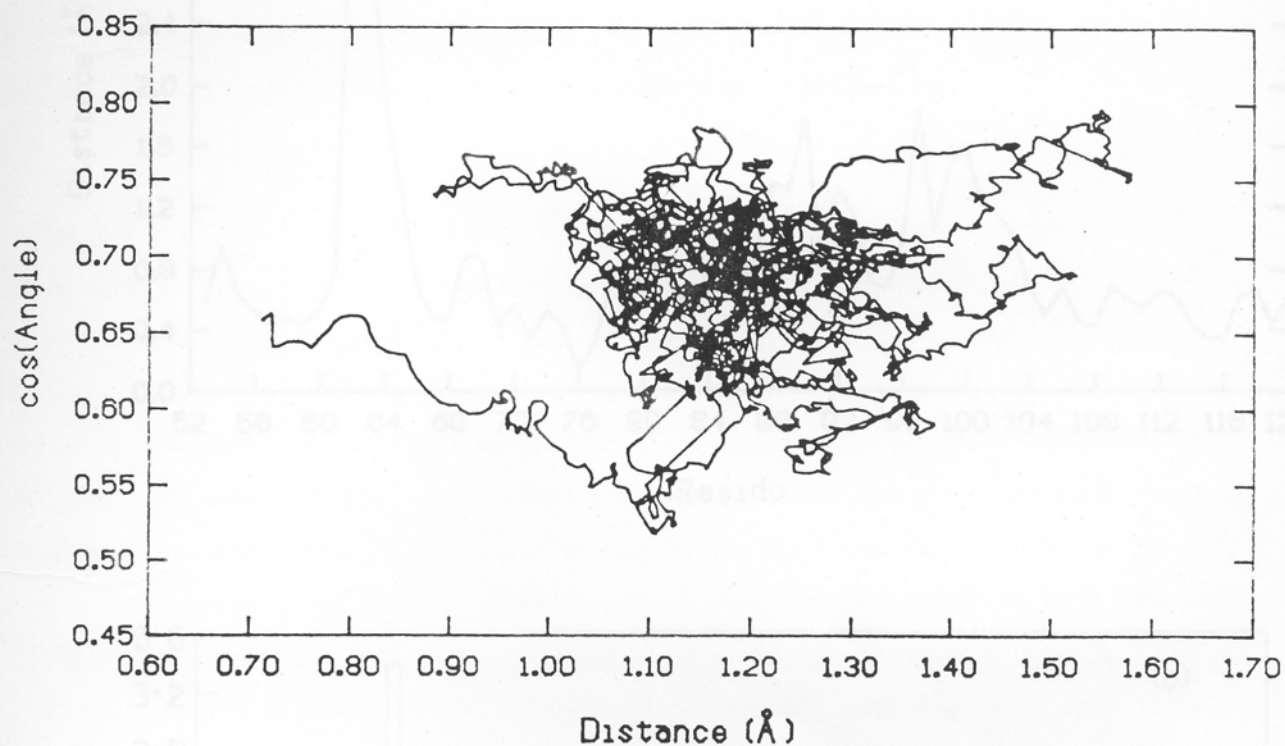


Figure III-3. Section de Poincaré de la trajectoire du CTF : la trajectoire du CTF est projetée sur un plan défini par deux variables globales indépendantes, la distance moyenne de la structure instantanée du CTF à sa structure cristallographique et le cosinus de l'angle que fait le vecteur correspondant à cette distance avec l'axe cartésien généralisé x .

Figure III-4. Distance entre la position moyenne et la position cristallographique de chacun des C_{α} du CTF : (a) trajectoire calculée avec CHARMM, (b) trajectoire simulée avec CROMOS (J. Aguilé & coll., 1993; reproduction faite avec l'autorisation du Journal of Molecular Biology).

atomes $C\alpha$:

$$d(t) = \sqrt{(\mathbf{d}(t))^2} = \sqrt{\frac{1}{N_\alpha} \sum_{i=1}^{N_\alpha} (\mathbf{r}_i(t) - \mathbf{r}_i(rX))^2}$$

la seconde est un cosinus défini ainsi :

$$\cos(t) = \frac{1}{d(t)\sqrt{N_\alpha}} \sum_{i=1}^{N_\alpha} (x_i(t) - x_i(rX))$$

Il s'agit du cosinus de l'angle que font le vecteur distance moyenne, $\mathbf{d}(t)$, et un vecteur à $3N$ dimensions représentant l'axe x du repère cartésien dans l'espace des configurations. Cette deuxième coordonnée décrit donc une projection sur l'axe x généralisé du vecteur dont la norme est $d(t)$, c'est-à-dire l'orientation de ce vecteur dans l'espace des configurations. La figure III-3 montre l'évolution concomitante de ces deux coordonnées générales entre $t=0$ et $t=65$ psec. Ce mode de représentation original, analogue aux sections de Poincaré, permet de visualiser ce qu'on peut appeler un régime dynamique stationnaire pour un système aussi complexe qu'une protéine. La plupart du temps, le système évolue au voisinage d'un point bien précis du plan défini par les deux coordonnées décrites ci-dessus. Ce point est donc ce qu'on appelle un attracteur.

a) Comparaison de la structure moyenne et de la structure cristallographique.

i) Positions moyennes.

La figure III-4a montre, pour chacun des $C\alpha$, la distance entre la position moyenne de l'atome lors de la simulation et sa position cristallographique. Ces distances sont supérieures à 1,2 Å pour l'atome $C\alpha$ de certains résidus des boucles reliant les structures secondaires βA et αA (résidus 62 à 65), αB et βB (résidus 88 à 90), βB et αC (résidus 97, 99 et 100). Pour la plupart des autres résidus, elles sont inférieures à 0,8 Å. L'atome $C\alpha$ des résidus 62-63 de la boucle βA - αA est à environ 3,5 Å de sa position cristallographique, résultat tout à fait similaire à celui trouvé par Åqvist (Fig. III-4b). La cause du déplacement de la position moyenne de cette boucle doit être celle avancée par cet auteur : un contre-ion et une molécule d'eau, présents dans des sites spécifiques près de cette boucle dans le cristal, sont absents de nos simulations. L'atome $C\alpha$ des résidus

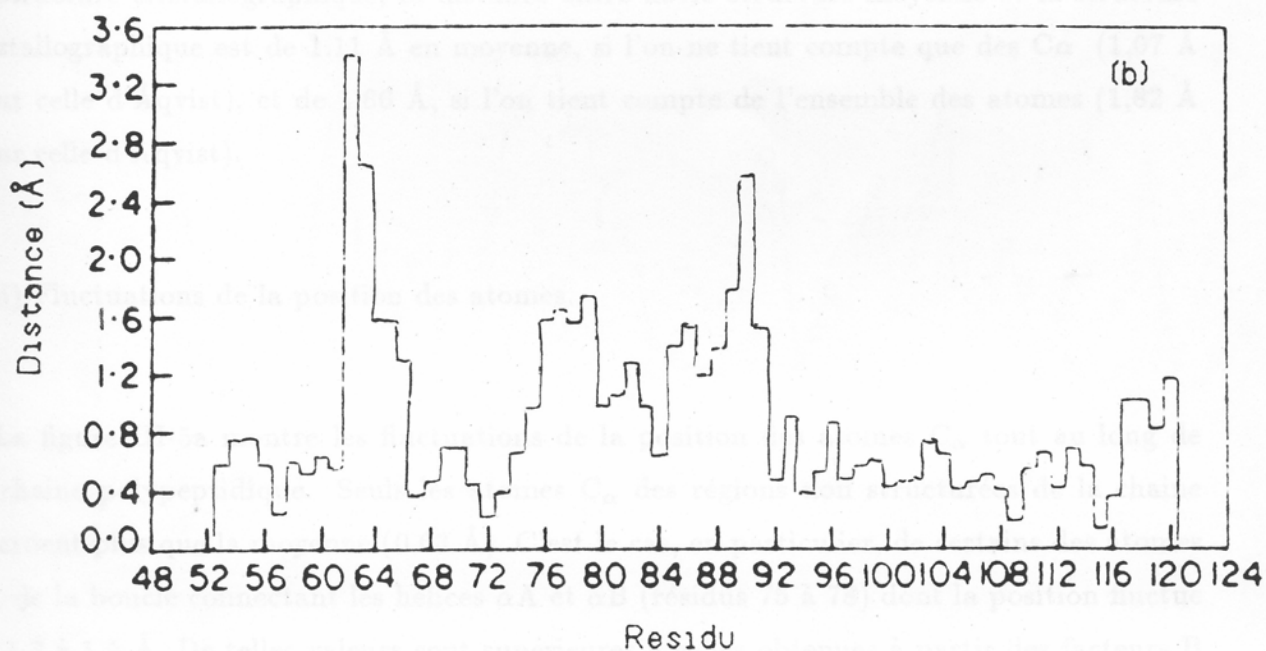
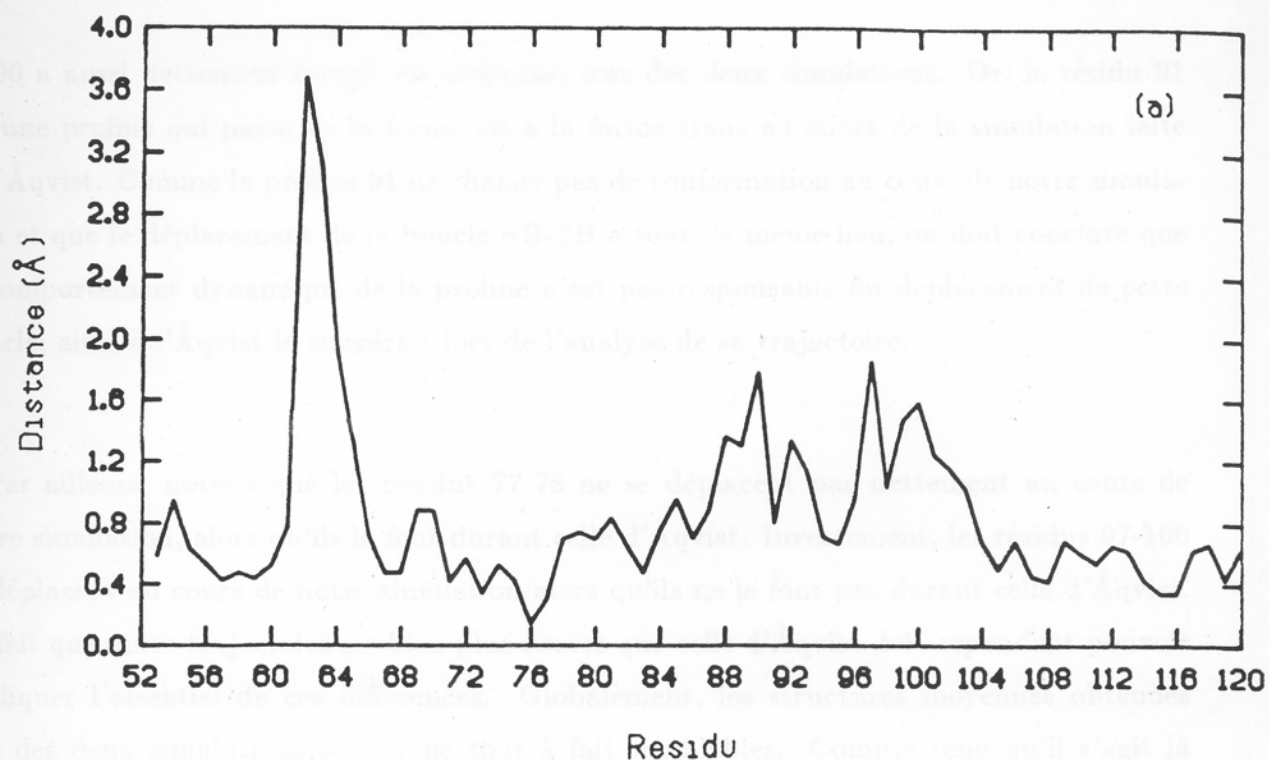


Figure III-4. Distance entre la position moyenne et la position cristallographique de chacun des C_{α} du CTF : (a) trajectoire calculée avec CHARMM. (b) trajectoire calculée avec GROMOS (J.Aqvist & col., 1985; reproduction faite avec l'autorisation du *Journal of Molecular Biology*).

88-90 a aussi nettement bougé, en moyenne, lors des deux simulations. Or, le résidu 91 est une proline qui passe de la forme cis à la forme trans au cours de la simulation faite par Åqvist. Comme la proline 91 ne change pas de conformation au cours de notre simulation et que le déplacement de la boucle α B- β B a tout de même lieu, on doit conclure que le comportement dynamique de la proline n'est pas responsable du déplacement de cette boucle, ainsi qu'Åqvist le suggérait lors de l'analyse de sa trajectoire.

Par ailleurs, notons que les résidus 77-78 ne se déplacent pas nettement au cours de notre simulation, alors qu'ils le font durant celle d'Åqvist. Inversement, les résidus 97-100 se déplacent au cours de notre simulation alors qu'ils ne le font pas durant celle d'Åqvist. Le fait que notre trajectoire est bien plus courte que celle d'Åqvist doit cependant pouvoir expliquer l'essentiel de ces différences. Globalement, les structures moyennes obtenues lors des deux simulations sont donc tout à fait semblables. Compte tenu qu'il s'agit là de simulations réalisées dans le vide, elles sont toutes deux remarquablement proches de la structure cristallographique; la distance entre notre structure moyenne et la structure cristallographique est de 1.11 Å en moyenne, si l'on ne tient compte que des C_α (1,07 Å pour celle d'Åqvist), et de 1,66 Å, si l'on tient compte de l'ensemble des atomes (1,82 Å pour celle d'Åqvist).

ii) Fluctuations de la position des atomes.

La figure III-5a montre les fluctuations de la position des atomes C_α tout au long de la chaîne polypeptidique. Seuls les atomes C_α des régions non structurées de la chaîne fluctuent plus que la moyenne (0.63 Å). C'est le cas, en particulier, de certains des atomes C_α de la boucle connectant les hélices α A et α B (résidus 75 à 78) dont la position fluctue de 1,2 à 1,5 Å. De telles valeurs sont supérieures à celles obtenues à partir des facteurs B expérimentaux (figure III-5b). Cela peut sembler évident, puisqu'une partie de cette boucle forme dans le cristal un brin β avec l'autre unité du dimère. Notons que la position moyenne de cette boucle est très voisine dans notre simulation de sa position cristallographique alors

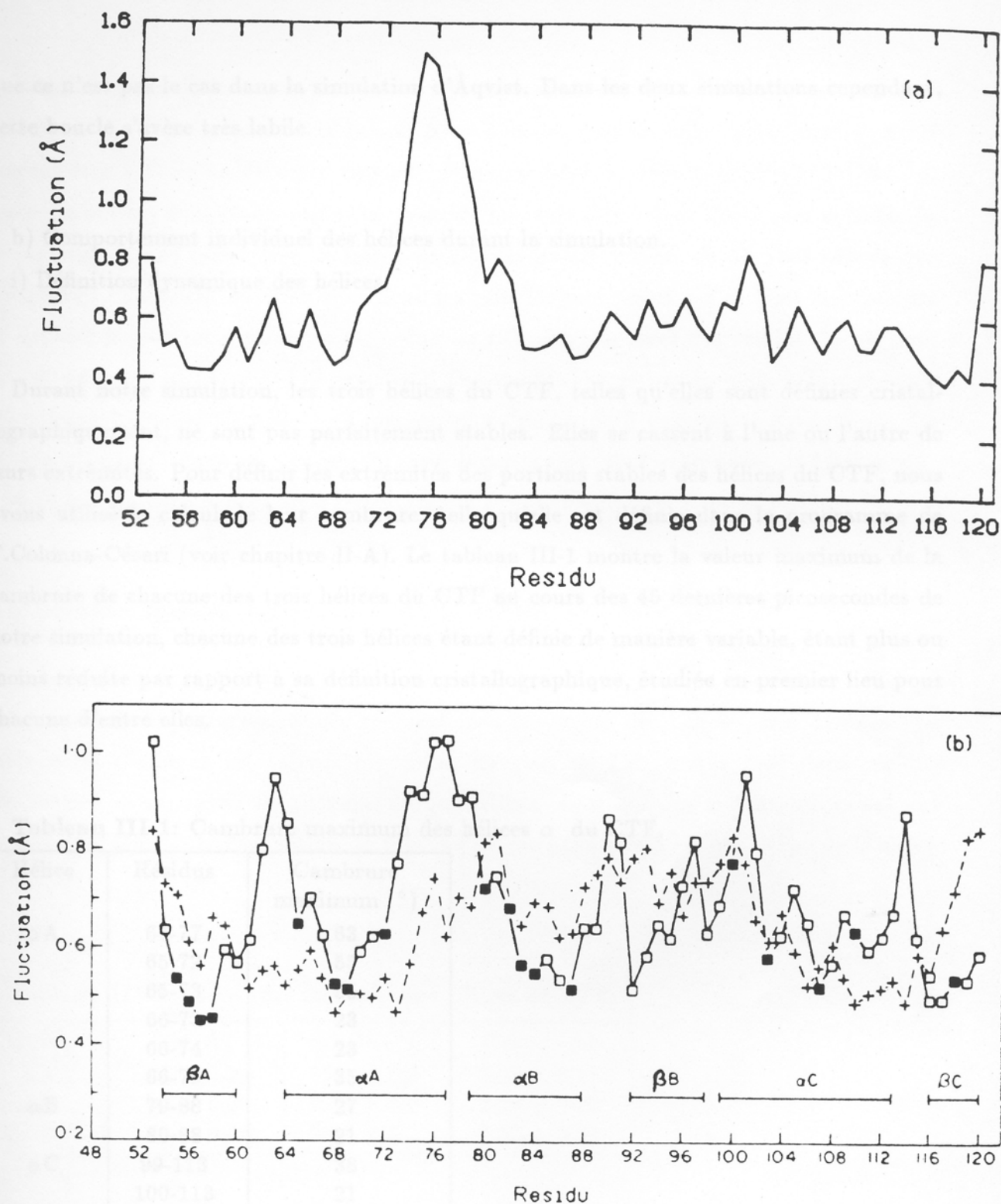


Figure III-5. Fluctuation de la position de chacun des atomes C_α du CTF au cours de la trajectoire : (a) trajectoire calculée avec CHARMM. (b) trajectoire calculée avec GROMOS (carrés et trait continu; les carrés sont noirs lorsque la surface accessible du résidu correspondant est inférieure à 1 \AA^2). Les facteurs B cristallographiques sont aussi indiqués (croix et tirets). Cette figure (J.Aqvist & col., 1985) est reproduite avec l'autorisation du *Journal of Molecular Biology*.

que ce n'est pas le cas dans la simulation d'Åqvist. Dans les deux simulations cependant, cette boucle s'avère très labile.

b) Comportement individuel des hélices durant la simulation.

i) Définition dynamique des hélices.

Durant notre simulation, les trois hélices du CTF, telles qu'elles sont définies cristallographiquement, ne sont pas parfaitement stables. Elles se cassent à l'une ou l'autre de leurs extrémités. Pour définir les extrémités des portions stables des hélices du CTF, nous avons utilisé le calcul de leur cambrure, telle qu'elle est définie dans le programme de F.Colonna-Césari (voir chapitre II-A). Le tableau III-1 montre la valeur maximum de la cambrure de chacune des trois hélices du CTF au cours des 45 dernières picosecondes de notre simulation, chacune des trois hélices étant définie de manière variable, étant plus ou moins réduite par rapport à sa définition cristallographique, étudiée en premier lieu pour chacune d'entre elles.

Tableau III-1: Cambrure maximum des hélices α du CTF.

Hélice	Résidus	Cambrure maximum (°)
αA	64-77	63
	65-72	55
	65-73	55
	66-73	23
	66-74	23
αB	66-75	35
	79-88	27
	80-88	21
αC	99-113	38
	100-113	21

Les trois hélices sont suffisamment stables et régulières pour permettre de calculer leur

On remarque notamment que l'hélice αA , telle qu'elle est définie cristallographiquement,

se casse au cours de la simulation (cambrure maximum de 63°). Il faut donc réviser sa définition. Le déplacement de la boucle $\beta A-\alpha A$ et les mouvements de la boucle $\alpha A-\alpha B$ doivent être à l'origine de cette cassure. Pour modifier la définition de l'hélice αA , et des deux autres, nous avons utilisé la méthode suivante : la valeur maximum normale de la cambrure d'une hélice lors de notre simulation paraît voisine de 25° ; les hélices que nous avons finalement étudiées sont donc les plus longues parmi celles dont la cambrure ne dépasse jamais cette valeur, c'est-à-dire les hélices composées des résidus 66-74, pour αA , 80-88 pour αB , et 100-113 pour αC . Ces modifications par rapport à la définition cristallographique de ces hélices ne concernent qu'un résidu à l'extrémité N-terminale des hélices αB et αC , mais en concernent cinq répartis aux deux extrémités de l'hélice αA . Notons que les trois hélices ainsi redéfinies ont respectivement 9, 9, et 14 résidus.

ii) Structure moyenne des hélices.

On a étudié les caractéristiques géométriques moyennes des hélices ainsi définies. Le tableau III-2 montre les valeurs moyennes du pas, de l'angle que font deux résidus consécutifs, et du rayon de chacune de ces hélices, ainsi que les valeurs moyennes des erreurs affectant la détermination de chacune de ces quantités à chaque étape de la simulation.

Tableau III-2 : Structure moyenne des hélices du CTF.

Hélice	Pas (\AA)	Angle ($^\circ$)	Rayon (\AA)
αA	$1,61 \pm 0,27$	$99,5 \pm 9,0$	$2,25 \pm 0,26$
αB	$1,52 \pm 0,20$	$98,8 \pm 5,9$	$2,29 \pm 0,20$
αC	$1,50 \pm 0,23$	$99,3 \pm 8,7$	$2,32 \pm 0,24$

Les trois hélices sont suffisamment stables et régulières pour permettre de calculer leur axe sans ambiguïté en chaque point de la trajectoire. Cependant, il apparaît que le pas moyen de l'hélice αA a une valeur significativement différente de celle du pas d'une hélice

α idéale (1,5 Å). C'est aussi pour cette hélice que les trois paramètres géométriques de l'hélice sont le moins bien déterminés. Cette hélice est donc la moins régulière des trois. Les deux autres hélices ont une structure moyenne très proche de celle d'une hélice α idéale.

c) Comportement collectif des hélices.

La direction de l'axe de chacune des hélices est déterminée pour chacune des étapes de la simulation, à partir de la position des seuls atomes C_α . L'évolution temporelle des angles que font ces axes pris deux à deux, la fonction d'autocorrélation correspondante, calculée par une transformée de Fourier inverse (R.P.Futrelle & D.J.McGinty, 1971), et le spectre de puissance relative correspondant, calculé par une transformation de Fourier, sont montrés sur les figures III-6 à III-8.

i) Hélices αA et αC (figures III-6a à III-6c).

Un mouvement dont la fréquence moyenne est basse paraît se développer entre $t=0$ et $t=45$ psec, environ. Sa présence est, entre autres, suggérée par la lente diminution de la valeur de la fonction d'autocorrélation. Ensuite, le comportement dynamique du CTF, tel qu'il est visualisé par l'évolution de l'angle que font les deux hélices αA et αC , paraît changer. Dans le spectre de puissance, la plus basse des fréquences, 1 cm^{-1} , rend compte du changement de régime dynamique qui se produit vers $t=45$ psec.

ii) Hélices αA et αB (figures III-7a à III-7c).

Le spectre de puissance de l'évolution en fonction du temps de la valeur de l'angle que font les hélices αA et αB est plus riche que le précédent. Un événement brutal affecte aussi

Figure III-6. Mouvement relatif des hélices αA et αC : a) évolution de la valeur de l'angle que font les axes de ces deux hélices au cours de la trajectoire du CTF; b) fonction d'autocorrélation correspondante; c) Spectre de puissance relative correspondant.

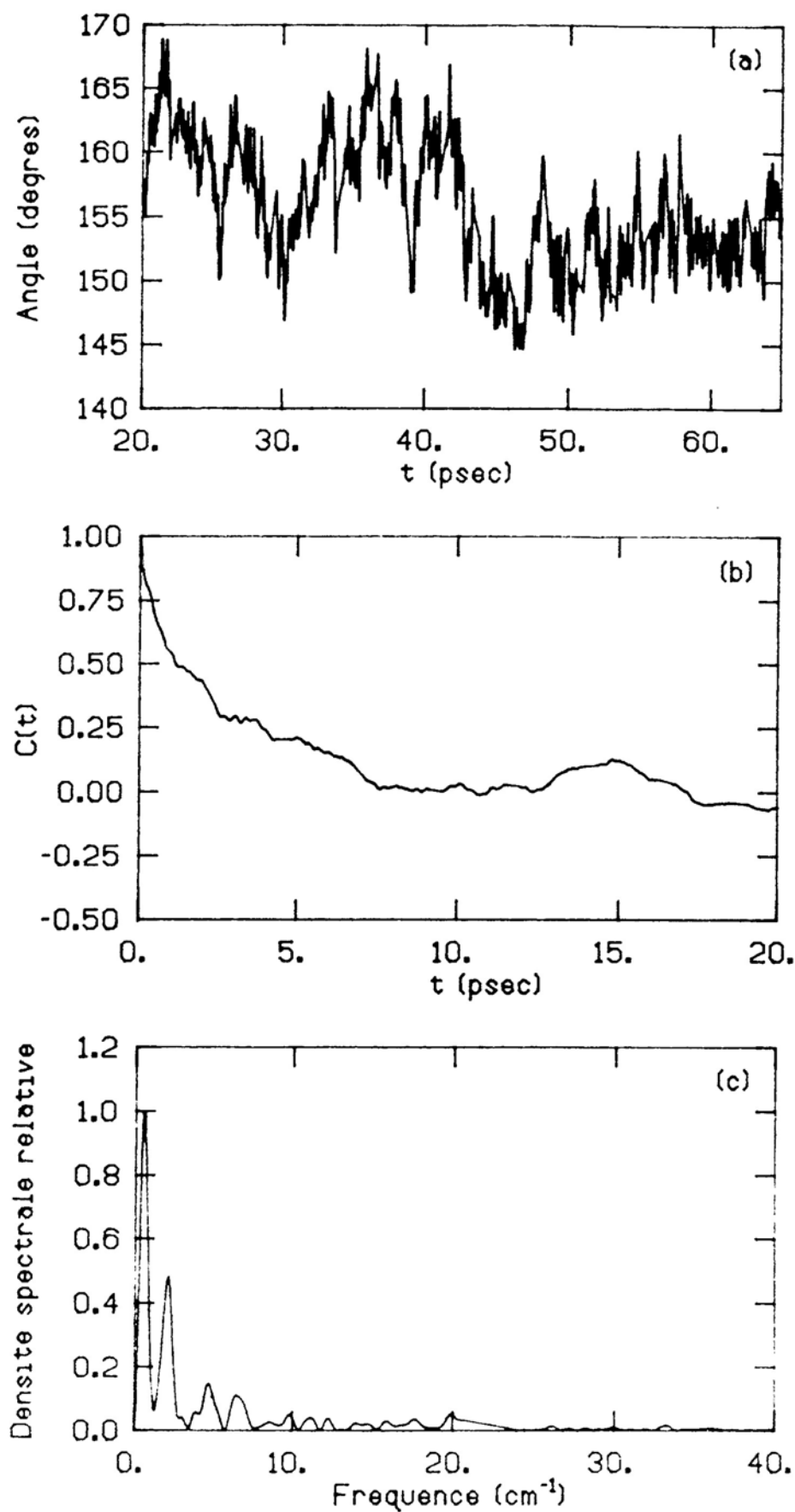


Figure III-6. Mouvement relatif des hélices αA et αC : a) évolution de la valeur de l'angle que font les axes de ces deux hélices au cours de la trajectoire du CTF; b) fonction d'autocorrélation correspondante; c) Spectre de puissance relative correspondant.

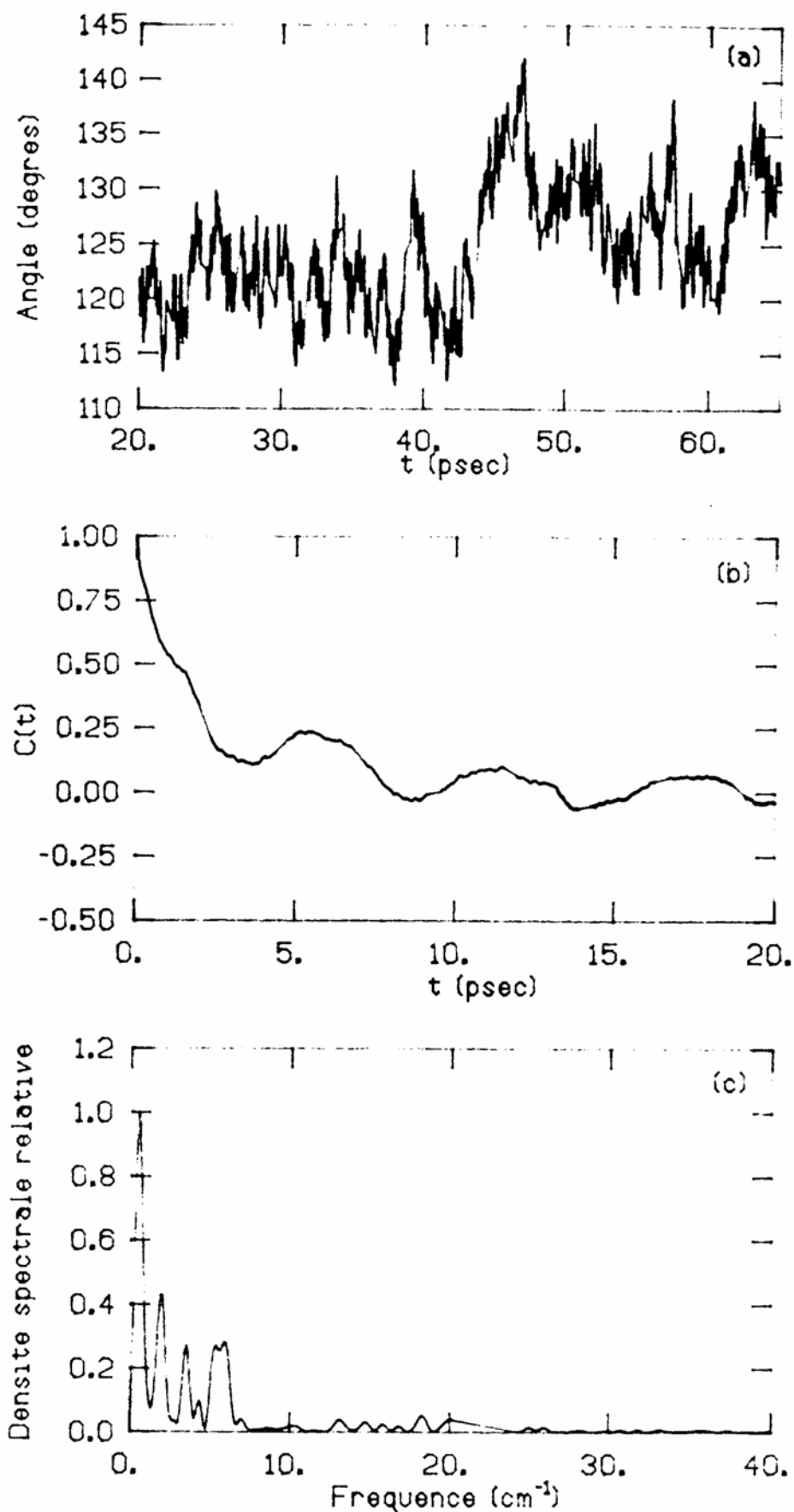


Figure III-7. Mouvement relatif des hélices αA et αB : a) évolution de la valeur de l'angle que font les axes de ces deux hélices au cours de la trajectoire du CTF; b) fonction d'autocorrélation correspondante; c) Spectre de puissance relative correspondant.

l'évolution de cet angle vers $t=45$ psec. L'analyse détaillée du comportement dynamique de l'hélice αA montre que vers cet instant ses paramètres géométriques atteignent des valeurs extrêmes : 1,39 Å pour le pas, $94,5^\circ$ pour l'angle, et 2,45 Å pour le rayon, mais que par la suite ses caractéristiques moyennes ne sont pas radicalement modifiées. Le changement de régime dynamique observé doit donc être le symptôme d'un changement de conformation du CTF entraînant une modification de la position relative de l'hélice αA par rapport aux deux autres.

iii) Hélices αB et αC (figures III-8a à III-8c).

La corrélation entre les mouvements des hélices αB et αC décrite par Åqvist (figure III-9) est nettement retrouvée, bien que notre trajectoire soit plus courte la sienne. Un mouvement de basse fréquence, 5 cm^{-1} environ, et d'une amplitude d'environ 10° domine le spectre du déplacement relatif des axes de ces deux hélices, alors que la plus courte distance entre ces deux axes est, en moyenne, proche de 15 Å. On remarque que la fréquence de ce mouvement paraît varier au cours de la dynamique, à partir de $t=45$ ou $t=50$ psec, ce qui conforte l'hypothèse d'un changement majeur de conformation, c'est-à-dire d'une transition affectant le comportement dynamique du CTF dans son ensemble, au voisinage de $t=45$ psec.

Le mouvement quasi-périodique de l'hélice αB décrit par Åqvist est mis en évidence lorsque l'hélice αC est utilisée comme référence pour étudier son mouvement. Ceci résulte probablement du fait que l'hélice αC ne doit pas avoir la possibilité d'effectuer d'aussi grands mouvements que les hélices de la boucle αA - αB . Quoiqu'il en soit, l'analyse de notre simulation du CTF montre que les résultats obtenus par Åqvist avec le programme GROMOS sont reproductibles et, de plus, peu sensibles à des modifications du modèle et de la méthode employés pour les obtenir.

Cependant, le mouvement périodique relatif de basse fréquence des hélices αB et αC

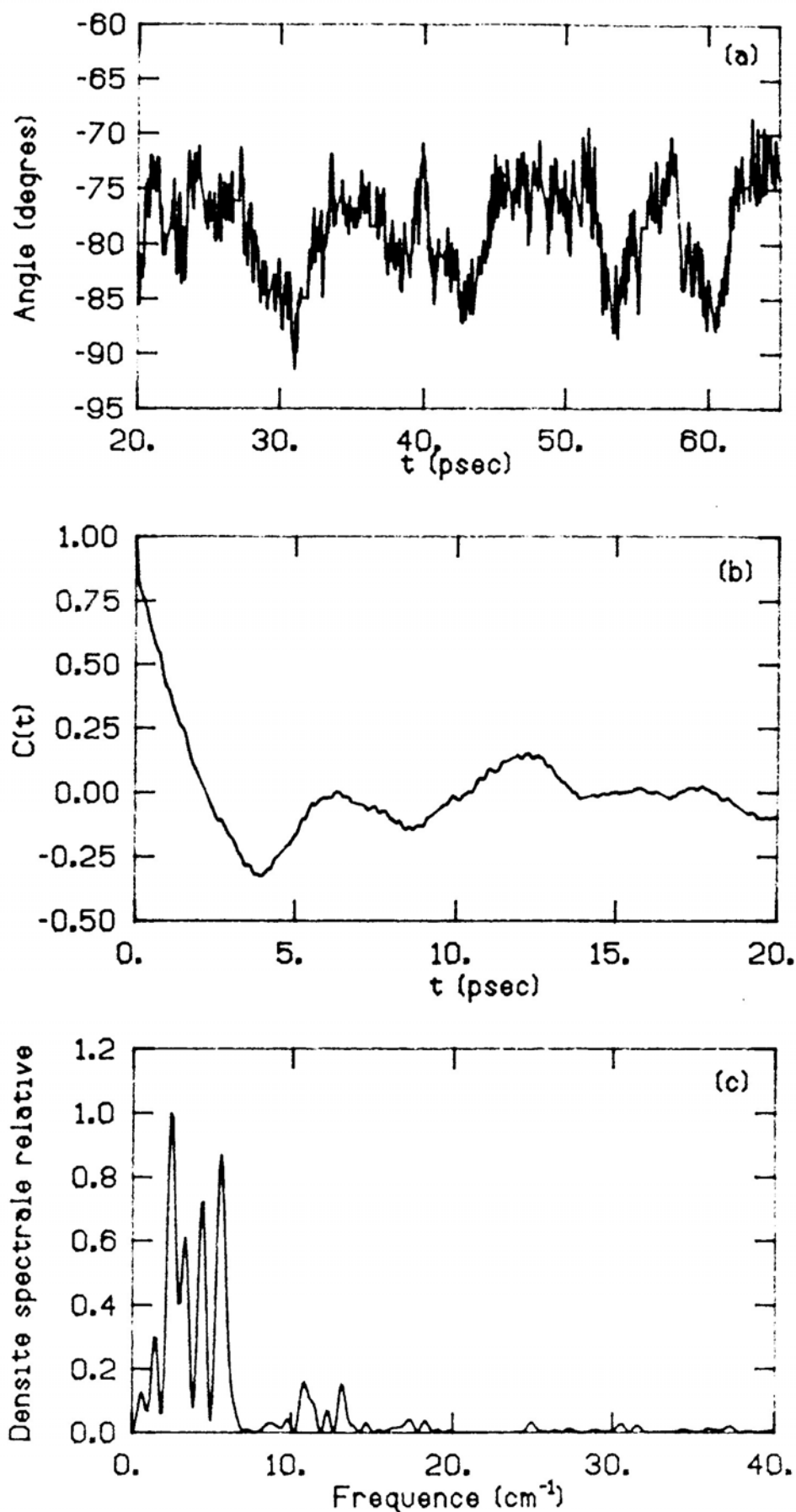


Figure III-8. Mouvement relatif des hélices αB et αC : a) évolution de la valeur de l'angle que font les axes de ces deux hélices au cours de la trajectoire du CTF; b) fonction d'autocorrélation correspondante; c) Spectre de puissance relative correspondant.

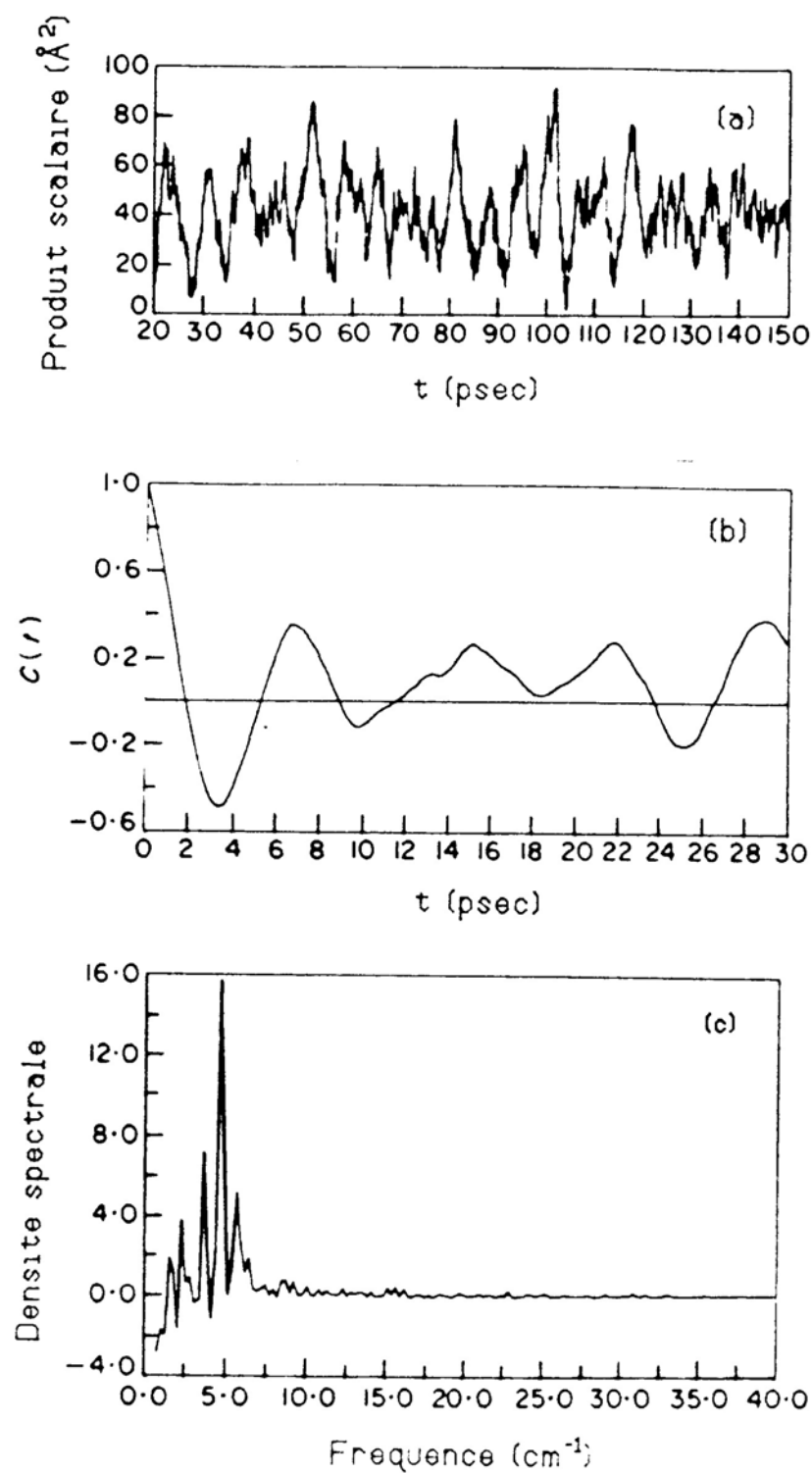


Figure III-9. Mouvement relatif des hélices αB et αC (GROMOS): a) évolution de la valeur de l'angle que font les axes de ces deux hélices au cours de la trajectoire du CTF calculée avec le programme GROMOS; b) fonction d'autocorrélation correspondante; c) Spectre de puissance relative correspondant (J.Aqvist & col., 1985; ces trois reproductions sont faites avec l'autorisation du *Journal of Molecular Biology*).

a été mis en évidence lors de simulations qui toutes ont été obtenues avec des conditions de calcul des termes électrostatiques de la fonction énergie potentielle très similaires. Or, Smith a montré que la densité des fréquences des modes de vibration, $g(w)$, est très sensible, dans le domaine des basses fréquences, à la méthode employée pour calculer le terme électrostatique de la fonction énergie potentielle (J.Smith & col., 1989). On peut donc se demander si le mouvement périodique relatif des hélices αB et αC est, ou non, sensible aux conditions particulières du calcul des interactions électrostatiques, c'est-à-dire s'il est, ou non, un artefact d'origine méthodologique.

Dans un premier temps, nous allons vérifier que dans notre cas la densité des fréquences des modes de vibration du CTF, $g(w)$, est peu sensible à la méthode employée pour calculer le terme électrostatique de la fonction énergie potentielle. La figure III-10 montre le $g(w)$ du CTF, calculé à partir des fréquences propres de la structure initiale de notre première simulation; notons que quelques basses fréquences propres ont une valeur de l'ordre de grandeur de celles qui dominent le spectre du mouvement relatif des hélices αB et αC . Par exemple, la plus basse des fréquences propres de cette structure est de $4,8 \text{ cm}^{-1}$.

Pour montrer que le mouvement relatif des hélices αB et αC n'est pas un artefact d'origine méthodologique, et tenter ensuite de comprendre son origine physique, il nous faut aussi vérifier qu'il existe lors d'une simulation durant laquelle le calcul du terme électrostatique est très différent. Pour ce faire, nous avons choisi de réaliser une trajectoire "exacte", c'est-à-dire une trajectoire durant laquelle toutes les interactions entre atomes sont prises en compte. Ainsi, à chaque étape de cette nouvelle simulation, 174796 distances sont calculées, au lieu d'environ 45000 lorsque seuls les atomes situés à moins de 10.5 \AA les uns des autres sont pris en compte.

E. Simulation incluant les effets électrostatiques à longue distance.

Dans un premier temps, afin d'estimer l'influence de la modification du calcul du terme

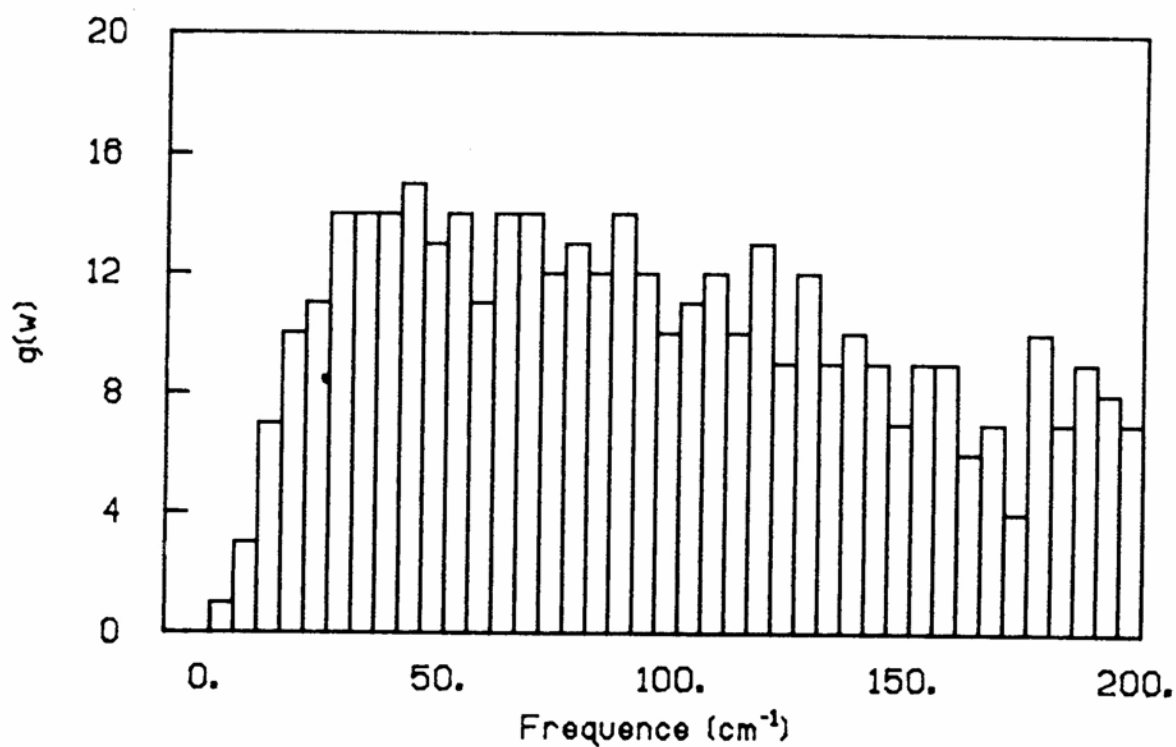


Figure III-10. Densité des fréquences des modes de vibration du CTF dans la gamme 0-200 cm^{-1} . L'énergie potentielle du CTF a été minimisée à partir de la structure cristallographique, avec une distance de coupure fixée à 10.5 Å pour le calcul des interactions électrostatiques.

électrostatique de la fonction énergie potentielle sur l'allure du $g(w)$ du CTF, l'énergie potentielle de la structure initiale de la précédente simulation est minimisée avec les nouvelles conditions électrostatiques. Précédemment, l'énergie potentielle de cette structure était de -2380,9 Kcal/Mole. Avec les nouvelles conditions, l'énergie potentielle du CTF passe de -2954,5 à -2965,4 Kcal/Mole lors de la minimisation, mais sa structure n'évolue pas beaucoup. En effet, la distance moyenne entre les structures initiales et finales est de 0,44 Å. La figure III-11 montre la densité des fréquences des modes de vibration, $g(w)$, calculée pour la nouvelle structure. En comparant les figures III-10 et III-11, on ne constate que des différences mineures dans l'allure du $g(w)$ du CTF. Un effet du type de celui signalé par Smith, qui comparait les effets des formules de coupure de type SWITCH et SHIFT sur le $g(w)$ du BPTI, n'est donc pas retrouvé ici. Cependant, afin de conforter l'idée que le mouvement périodique relatif des hélices αB et αC du CTF est peu sensible aux détails du modèle et des méthodes employées, une simulation durant laquelle toutes les interactions électrostatiques sont prises en compte est tout de même réalisée. Après avoir confirmé que l'origine de ce mouvement n'est pas un artefact méthodologique dû au mode de calcul des contributions électrostatiques du potentiel du CTF, nous allons chercher à comprendre l'origine physique du mouvement relatif des hélices αB et αC .

1) Analyse de la trajectoire.

Dans le but de comparer nos deux trajectoires, et de réaliser une simulation qui soit indépendante de notre première étude, nous avons appliqué rigoureusement la même méthode que précédemment.

Tout d'abord, l'énergie potentielle de la structure cristallographique est minimisée. La structure obtenue à l'issue de cette minimisation est différente des précédentes; son énergie potentielle est de -2956,7 Kcal/Mole. La distance moyenne entre cette structure et la structure cristallographique est de 1,2 Å, lorsque tous les atomes sont pris en compte, et de 0,8 Å, lorsque seuls les C_α le sont. La densité des fréquences des modes de vibration,

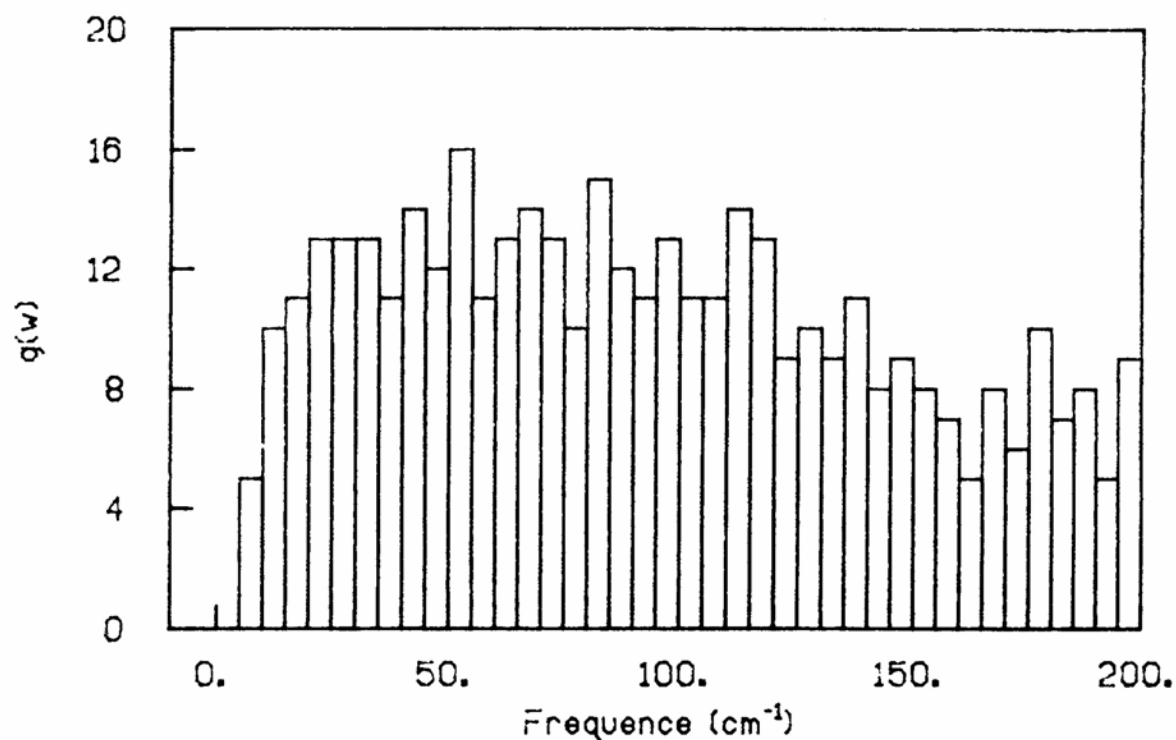


Figure III-11. Densité des fréquences des modes de vibration du CTF dans la gamme 0-200 cm^{-1} . L'énergie potentielle du CTF a été minimisée à partir de la structure cristallographique, avec une distance de coupure fixée à 10.5 Å pour le calcul des interactions électrostatiques puis minimisée de nouveau en tenant compte de toutes les interactions électrostatiques.

$g(w)$, de cette nouvelle structure ne diffère pratiquement pas de celle montrée sur la figure III-11. Un résultat similaire avait déjà été obtenu par Levitt pour le BPTI (M.Levitt & col., 1985). Le $g(w)$, qui est une quantité physiquement significative, ne paraît donc pas très sensible aux détails du repliement d'une protéine. Une telle conclusion peut plus particulièrement être tirée des résultats expérimentaux qui ont montré que le $g(w)$ obtenu par l'analyse de la diffusion inélastique des neutrons est peu différent pour des protéines totalement différentes, telles le BPTI, le lysozyme, la myoglobine (S.Cusack & col., 1988; J.Smith & col., 1989) et la Concalvaline-A (S.Cusack, 1990).

La figure III-12 montre l'évolution de la distance moyenne entre la structure instantanée et la structure cristallographique du CTF en fonction du temps, au cours de notre seconde simulation. Cette distance atteint lentement, en 30 psec environ, une valeur moyenne autour de laquelle elle oscille par la suite. Cette distance moyenne est de 1,4 Å, lorsque les seuls $C\alpha$ sont pris en compte, et de 2,2 Å lorsque tous les atomes sont inclus. Ces valeurs sont tout à fait similaires à celles obtenues par Åqvist, et un peu supérieures à celles obtenues précédemment.

La température du CTF, initialement de 540°K, descend en moins d'une picoseconde à 270°K, en moyenne, puis grimpe lentement pour fluctuer ensuite autour de 290°K, entre $t=30$ et $t=65$ psec. Déjà, entre $t=20$ et $t=25$ psec, elle a une valeur moyenne de 288°K. Contrairement au cas de la simulation précédente, la température ne dérive plus, en moyenne, après $t=30$ psec. Corrélativement, l'énergie totale du CTF ne dérive presque pas; la différence entre l'énergie totale du CTF à $t=0$ et à $t=65$ psec est de 0,4 Kcal/Mole, l'énergie totale étant en moyenne de $-2318,1 \pm 0,06$ Kcal/Mole, lors des dernières picosecondes de la simulation. L'origine de la dérive résiduelle de l'énergie totale doit provenir de l'algorithme SHAKE, la valeur tolérée pour les variations de la longueur des liaisons covalentes étant probablement un peu trop grande (0,001 Å).

La figure III-13 montre la représentation bidimensionnelle de la trajectoire du CTF entre

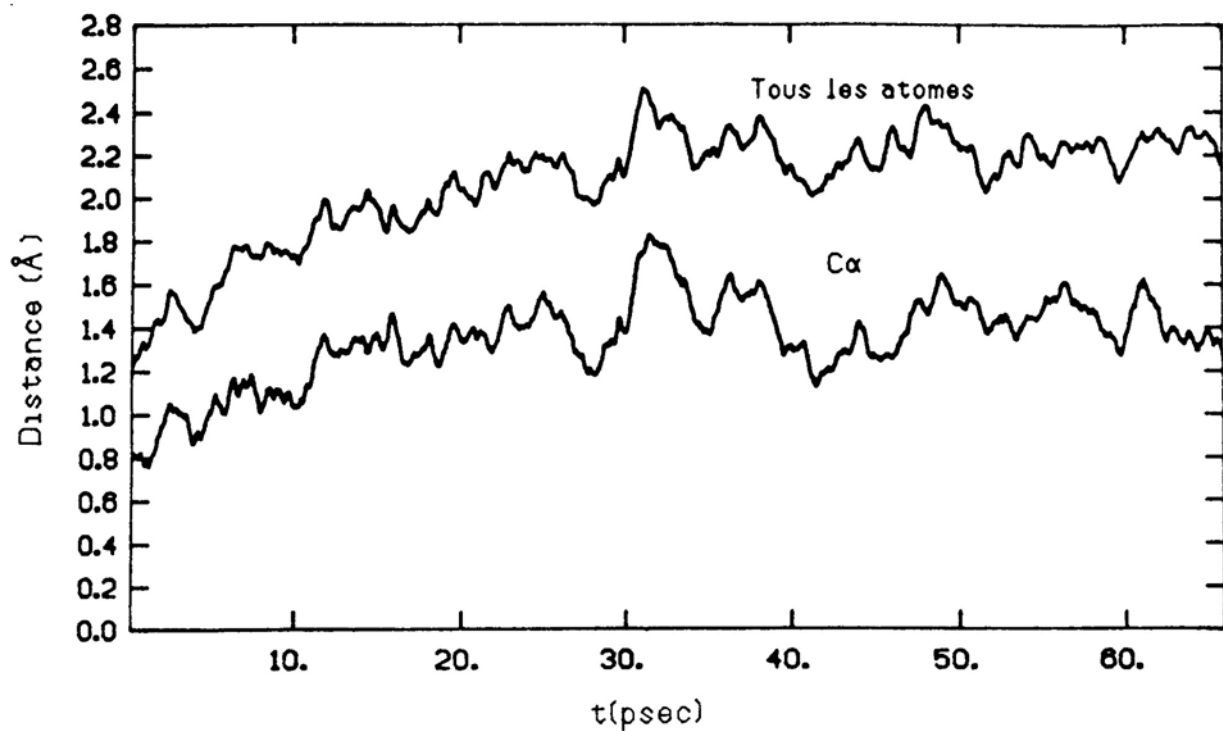


Figure III-12. Evolution en fonction du temps de la distance moyenne séparant la structure instantanée du CTF de sa structure cristallographique. Dans la courbe supérieure tous les atomes sont pris en compte lors du calcul de cette distance, alors que dans la courbe inférieure seuls les atomes C_{α} le sont.

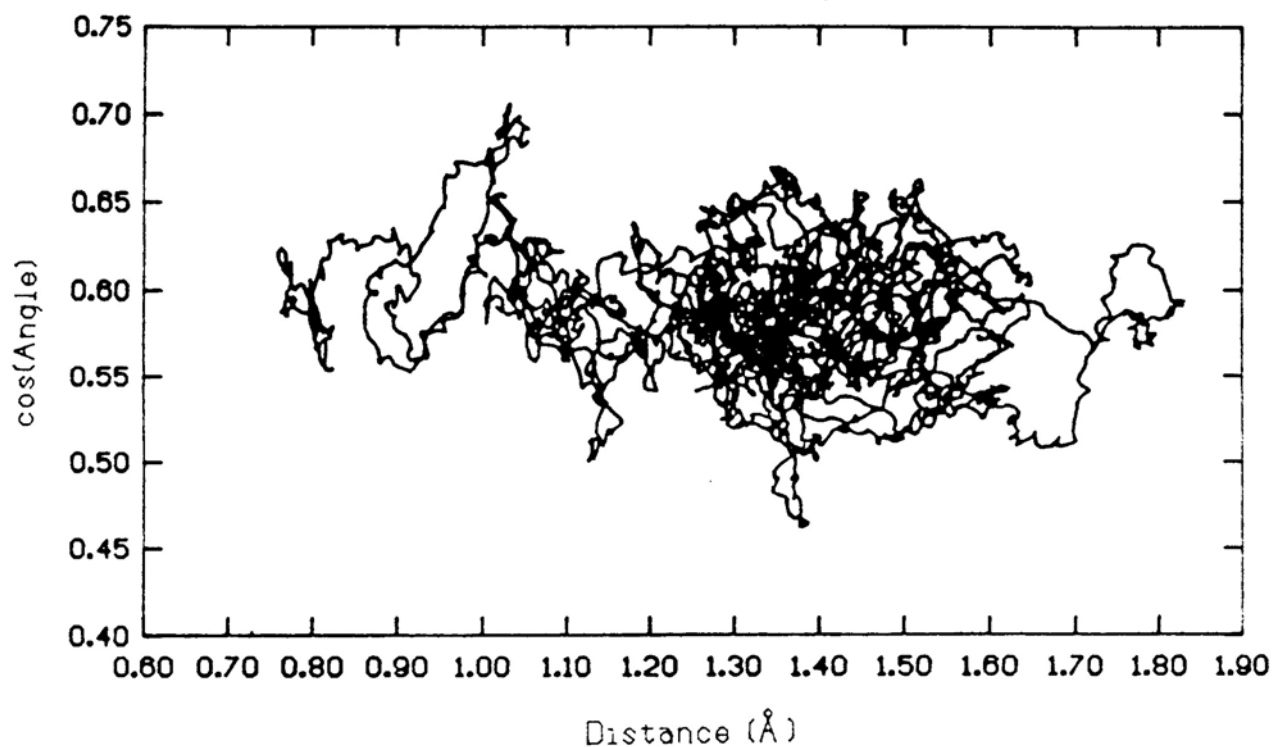


Figure III-13. Section de Poincaré de la trajectoire du CTF : La trajectoire du CTF est projetée sur un plan défini par deux variables globales indépendantes, la distance moyenne de la structure instantanée du CTF à sa structure cristallographique et le cosinus de l'angle que fait le vecteur distance correspondant avec l'axe cartésien généralisé x .

$t=0$ et $t=65$ psec. Dans un premier temps, le CTF rejoint un premier bassin d'attraction qu'il parcourt pendant quelques picosecondes; ensuite, il passe dans un second bassin dans lequel il se déplace durant l'essentiel de la simulation. Ce comportement explique la lente convergence de la valeur de la distance moyenne entre la structure instantanée et la structure cristallographique.

a) Comparaison de la structure moyenne et de la structure cristallographique.

i) Positions moyennes.

La distance entre la position moyenne de chacun des atomes C_α au cours de la simulation et sa position dans la structure cristallographique est montrée sur la figure III-14. Ces distances sont supérieures à 1,6 Å pour les atomes de certains résidus des boucles $\beta A-\alpha A$ et $\alpha A-\alpha B$. Le déplacement de la boucle $\beta A-\alpha A$ est donc tout aussi bien observé lors de nos deux simulations que lors de celle d'Åqvist. L'hypothèse donnée précédemment, concernant la cause de ce déplacement, s'en trouve renforcée, à savoir l'absence, dans toutes ces simulations, d'un contre-ion et d'une molécule d'eau "cristallographiques". Par contre, le déplacement de la boucle connectant les hélices αA et αB n'est observé que dans deux des trois simulations. Cependant, dans notre première simulation, durant laquelle ce déplacement ne se produit pas, la position des atomes de cette boucle fluctue beaucoup. L'instabilité de la position cristallographique des atomes de cette boucle est donc en fait observée dans toutes les simulations. Elle se manifeste soit par un déplacement de la position moyenne des atomes de cette boucle, soit par des fluctuations importantes de leur position. Dans le cristal, la position de cette boucle est stabilisée par les interactions entre les deux monomères du CTF. En effet, dans le dimère, la boucle $\alpha A-\alpha B$ d'un monomère est un brin du feuillet β de l'autre monomère. En conclusion, on peut dire que dans la présente simulation, toutes les différences importantes entre la structure dynamique moyenne et la structure cristallographique sont aisément rationalisables; ce sont celles que l'on aurait pu prévoir. Le reste de la structure n'est pas affecté de manière importante. En d'autres termes, les caractéristiques du repliement du CTF sont conservées au cours de toutes les trajectoires. Ce résultat signifie notamment que l'absence de solvant n'affecte

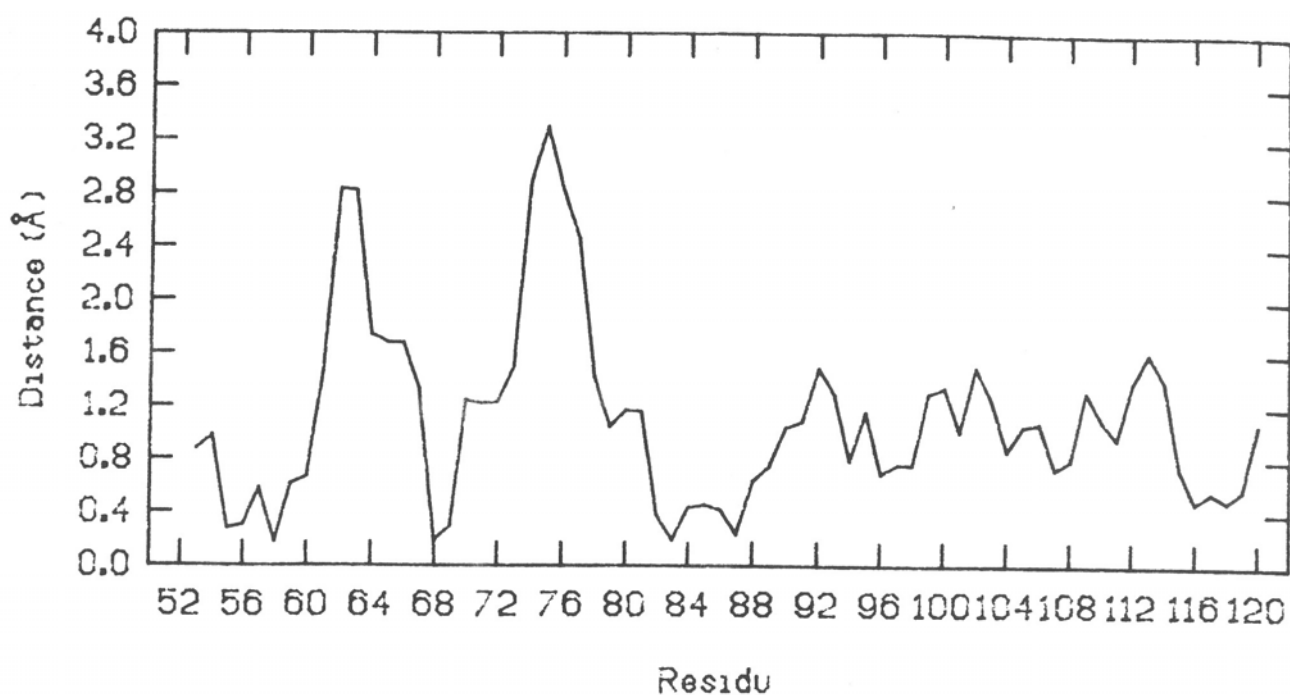


Figure III-14. Distance entre la position moyenne et la position cristallographique de chacun des C_{α} du CTF.

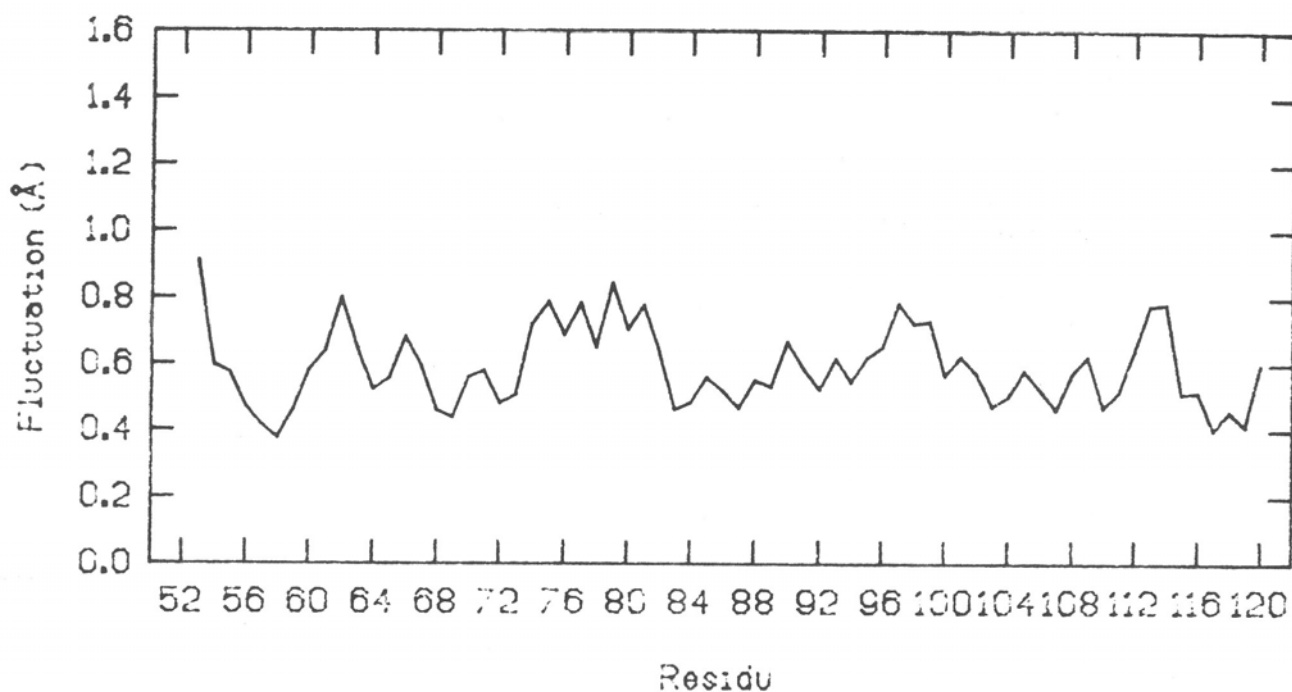


Figure III-15. Fluctuation de la position de chacun des atomes C_{α} du CTF au cours de la trajectoire.

pas la stabilité de la structure du CTF pendant le court laps de temps que durent ces simulations.

ii) Fluctuations des positions atomiques.

Les fluctuations de la position des atomes C_α sont montrées sur la figure III-15. Il n'y a aucun domaine de la séquence de la protéine dans lequel la position des atomes C_α fluctue de manière atypique, comme lors de la précédente simulation. Par ailleurs, l'ordre de grandeur de ces fluctuations est le même que celui obtenue à partir de l'analyse des facteurs B cristallographiques (voir la figure III-5b). Les propriétés dynamiques du CTF lors de cette simulation qui prend en compte les interactions électrostatiques à longue distance sont donc particulièrement satisfaisantes.

b) Propriétés moyennes des hélices.

Les propriétés moyennes des hélices du CTF au cours de la présente simulation sont très voisines de celles observées précédemment. Les valeurs maximales de la cambrure des hélices sont cependant un peu moins importantes : 17° pour l'hélice αA , 20° pour l' αB et 16° pour l' αC , le pas moyen de ces hélices étant, respectivement, de 1.57, 1.54 et 1.51 Å.

c) Mouvements relatifs des hélices.

Les figures III-16 et III-17 montrent l'évolution en fonction du temps de la valeur de l'angle que forment les axes des hélices αA et αB , d'une part, et αB et αC , d'autre part, ainsi que les fonctions d'autocorrélation et les spectres de puissance relative correspondants.

Le comportement dynamique relatif des hélices αB et αC est plus complexe que

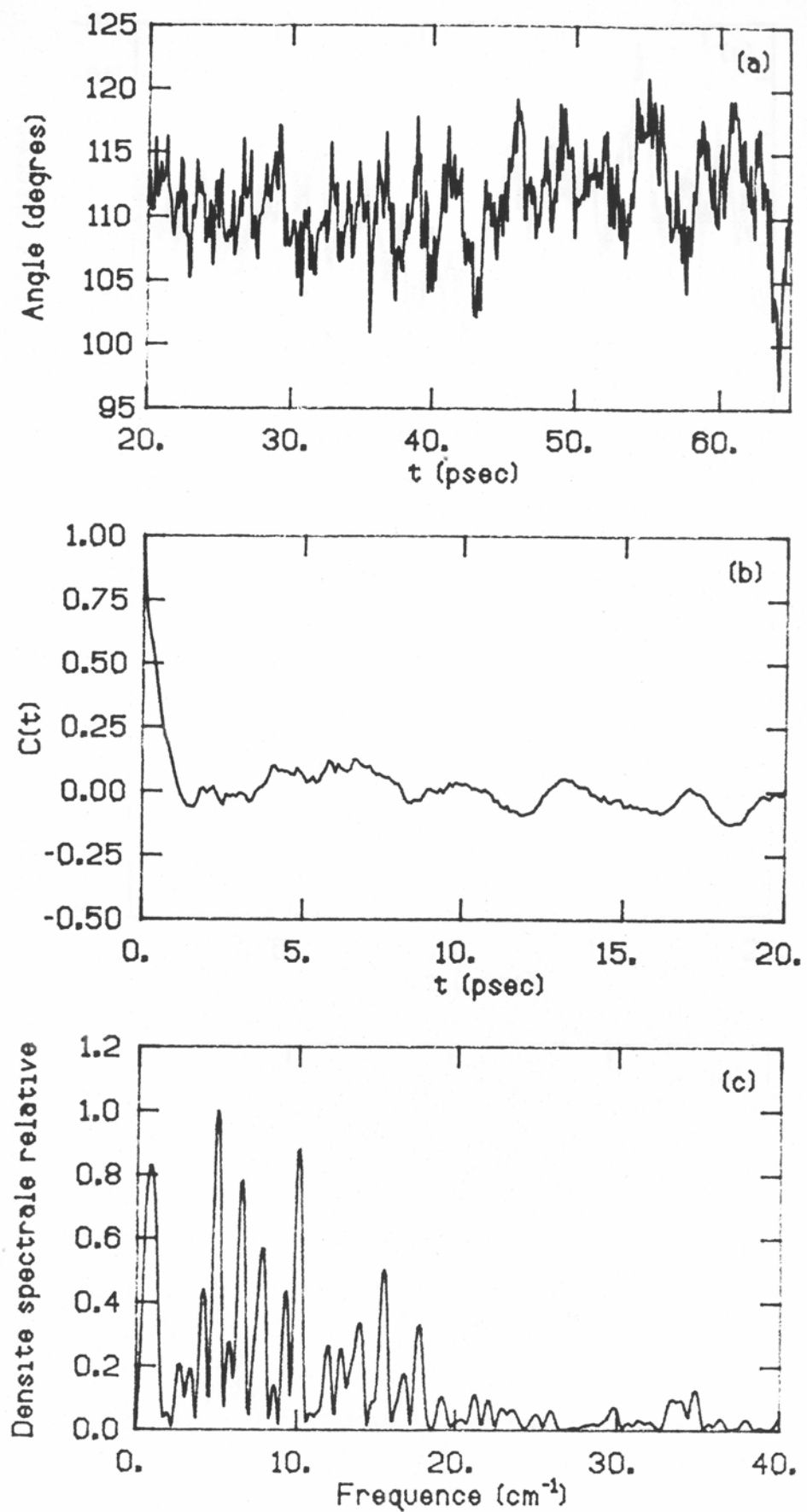


Figure III-16. Mouvement relatif des hélices αA et αB : a) évolution de la valeur de l'angle que font les axes de ces deux hélices au cours de la trajectoire du CTF; b) fonction d'autocorrélation correspondante; c) Spectre de puissance relative correspondant.

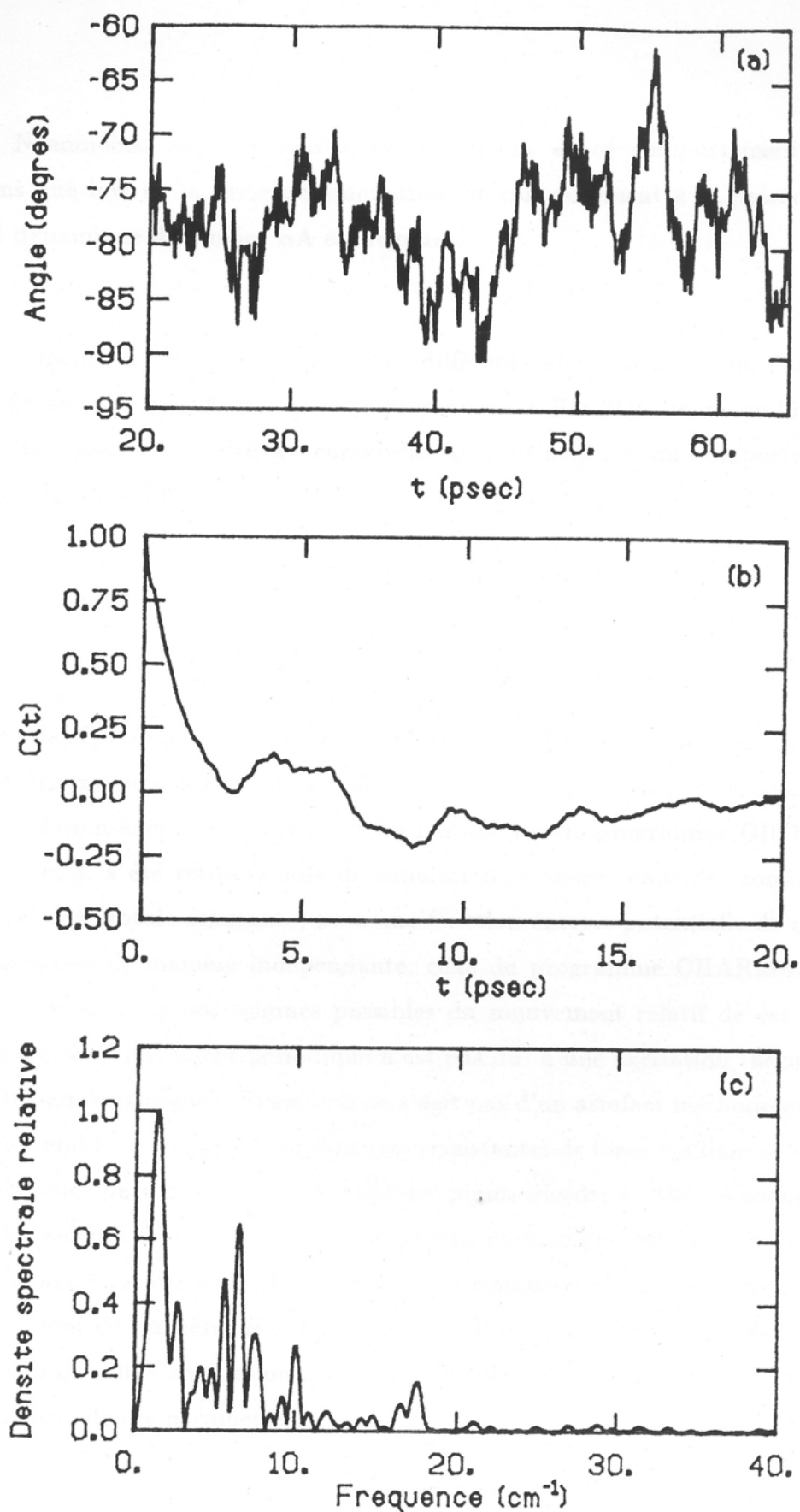


Figure III-17. Mouvement relatif des hélices αB et αC : a) évolution de la valeur de l'angle que font les axes de ces deux hélices au cours de la trajectoire du CTF; b) fonction d'autocorrélation correspondante; c) Spectre de puissance relative correspondant.

précédemment. Néanmoins, les caractéristiques principales de ce comportement sont présentes. Notons que lors de la présente simulation, et contrairement à la précédente, le comportement dynamique de l'hélice αA est régulier.

Pour résumer : même si nos deux trajectoires diffèrent sans doute autant l'une de l'autre que de celle calculée par Åqvist avec le programme GROMOS, leurs similitudes permettent de définir sans ambiguïté les caractéristiques principales du comportement dynamique intrinsèque du CTF.

d) Conclusion.

Le déplacement relatif périodique des hélices αB et αC du CTF, mis en évidence à la suite de simulations de dynamiques moléculaires réalisées dans des conditions de température constante avec la fonction énergie potentielle et les paramètres du programme GROMOS (J.Åqvist & coll., 1985), a été retrouvé lors de simulations réalisées dans des conditions microcanoniques (énergie totale constante) avec une fonction énergie potentielle de même forme mais paramétrisée de manière indépendante, celle du programme CHARMM. Ce résultat permet d'écarter deux des origines possibles du mouvement relatif de ces deux hélices. Tout d'abord, ce mouvement périodique n'est pas dû à une excitation thermique externe au CTF (le bain thermique). Ensuite, il ne s'agit pas d'un artefact méthodologique dû au choix d'un ensemble particulier de paramètres (constantes de force des liaisons chimiques, charges atomiques partielles, etc...). Ce dernier point, d'ordre méthodologique, est très important : les valeurs des paramètres des fonctions énergie potentielle sont affinées et modifiées chaque année; si certaines d'entre elles sont connues avec une précision satisfaisante, d'autres le sont de manière très approximative. Il importe donc de vérifier qu'un phénomène observé lors d'une simulation n'est pas sensible à une variation arbitraire et importante de la valeur de ces paramètres.

La persistance de cette propriété dynamique du CTF lorsqu'on modifie le modèle

Nous avons vérifié que le mouvement périodique relatif des hélices αB et αC du CTF

persiste lorsque le calcul de l'énergie électrostatique est fait très différemment. Le terme électrostatique est celui qui contribue le plus à la valeur de l'énergie potentielle d'une protéine; c'est aussi celui dont le calcul pose le plus de problèmes, à la fois pratiques et théoriques. Pratiquement, on ne peut, le plus souvent, calculer les interactions entre toutes les paires d'atomes d'une protéine; on ne tient alors compte que des atomes qui sont situés à une faible distance les uns des autres. Usuellement, l'ordre de grandeur de la distance maximum considérée (distance de coupure) est de 10 Å. Mais l'énergie potentielle d'interaction électrostatique est une quantité qui décroît lentement en fonction de la distance (en $\frac{1}{d}$): pour deux charges séparées de 5 Å, elle est de -66.4 Kcal/Mole, et elle est encore de -33.2 Kcal/Mole lorsque les deux charges sont séparées de 10 Å, ce qui n'est pas une valeur négligeable. Inversement, lorsqu'on calcule toutes les interactions électrostatiques d'une protéine, comme nous l'avons fait dans le cas de l'une de nos simulations du CTF, un effet physique est implicitement négligé, celui de la polarisation des nuages électroniques des atomes sous l'effet d'un champ électrostatique extérieur. Or, on sait que, pour des distances de séparation de 10 à 15 Å, cet effet modifie l'énergie d'interaction électrostatique d'environ 15 % (cet effet est une des composantes de la constante diélectrique). En son absence, l'incidence des interactions à longue distance sur le comportement dynamique d'une protéine est donc surestimée lorsque toutes les interactions atomiques sont prises en compte. Lors de notre première simulation, qui cherchait à reproduire l'essentiel des caractéristiques de celles d'Aqvist, une distance de coupure était introduite dans le calcul de l'énergie électrostatique. Les interactions à longue distance étaient donc négligées. En réalisant une seconde simulation, où elles sont surestimées, on a donc vérifié que le mouvement périodique des hélices B et C du CTF n'est pas un artefact méthodologique rendu possible par la présence d'une distance de coupure ou par l'absence de ces interactions à longue distance. Ce résultat doit avoir une valeur très générale, puisque Brooks a montré qu'avec la méthode SHIFT, que nous avons utilisée, la présence d'une distance de coupure affecte peu les résultats obtenus à la suite d'une simulation (R.J.Loncharich & B.Brooks, 1989).

La persistance de cette propriété dynamique du CTF lorsqu'on modifie le mode de calcul de son énergie potentielle d'interaction est un argument en faveur de la validité

de l'approche méthodologique mise en oeuvre lorsqu'on applique les méthodes de la dynamique moléculaire à l'étude d'une protéine. Ce résultat suggère que les propriétés dynamiques majeures d'une protéine peuvent être obtenues à partir d'un modèle *grosso modo* correct.

2) Calcul du mouvement des hélices par la théorie des modes normaux.

Pour tenter de comprendre l'origine physique du mouvement périodique relatif des hélices αB et αC , les modes normaux du CTF ont été analysés. Cependant, à partir des matrices A et Λ , calculées pour une structure donnée, on ne peut aisément obtenir analytiquement que des informations de nature locale : fluctuations des positions atomiques, des angles dièdres, des longueurs de liaison, etc... Il est plus difficile d'en extraire des informations de nature globale, telles l'évolution temporelle de l'angle que font deux hélices. C'est pourquoi nous avons calculé, à partir des équations (13) et (14), les trajectoires correspondant aux solutions analytiques données par la théorie des modes normaux. La température moyenne de ces trajectoires est de $290^\circ K$, comme lors de nos simulations de dynamique moléculaire. Ces trajectoires "harmoniques" ont été analysées de la même façon que les deux trajectoires étudiées précédemment.

a) Calcul classique.

L'énergie potentielle des structures du CTF à $t=0, 30, 40, 50, 60$ psec, est minimisée. En d'autres termes, à divers instants de notre dernière simulation, le CTF est brutalement refroidi. Les propriétés dynamiques des structures gelées ainsi obtenues sont étudiées. Pour cela, les trajectoires harmoniques, calculées à partir des modes normaux de chacune des cinq structures gelées, sont analysées et comparées.

Tout d'abord, on note que l'énergie potentielle de ces structures gelées est relativement

constante; elle varie de -3024 ($t=30$ psec) à -3032 Kcal/Mole ($t=50$ psec). Ce résultat, comparable à celui obtenu par Go lors d'une étude de la myoglobine (N.Go, 1989) est une indication supplémentaire en faveur de l'hypothèse que le CTF a un régime dynamique qui est stationnaire durant la période étudiée. Cependant, notons que les structures gelées obtenues sont toutes différentes. Notamment, la valeur de la plus basse fréquence propre de ces structures varie de 5,6 à 8,2 cm^{-1} , selon les cas. Par ailleurs, la température moyenne que l'on peut calculer à partir des valeurs minimales de l'énergie potentielle dans l'attracteur que parcourt le CTF est d'environ 300°K, soit à peu près 10°K de plus que la température moyenne du CTF lors de notre simulation; une partie de l'énergie potentielle n'est donc pas convertie en énergie cinétique. En d'autres termes, le CTF ne passe jamais par le fond des puits de la surface d'énergie potentielle qu'il parcourt. L'allure de cette surface étant très accidentée, ce phénomène n'est pas étonnant; si les chemins qui conduisent au fond des puits sont complexes et peu nombreux, la probabilité pour que le système atteigne l'un deux au cours d'une simulation de courte durée est extrêmement faible. Ce phénomène indique aussi que les structures gelées ne sont pas forcément idéales pour étudier les caractéristiques dynamiques du CTF à température ordinaire.

Les comportements dynamiques des structures gelées sont très semblables. Les fluctuations de la position des atomes C_α varient de 0,38 à 0,41 Å, selon les trajectoires, valeurs très inférieures à celles observées lors des trajectoires de dynamique moléculaire. Par ailleurs, l'angle moyen que font les hélices αB et αC varie de -78° à -80° tandis que celui fait par les hélices αA et αB varie de 108° à 112°. Les fluctuations de la valeur de ces deux angles sont comprises entre 2.8 et 3.1°, selon les paires d'hélices et les trajectoires. Quand aux fréquences qui dominent ces mouvements relatifs, elles sont élevées : de 8 à 10 cm^{-1} pour le mouvement relatif des hélices αB et αC . De telles fréquences ne peuvent rendre compte du phénomène observé lors des simulations de dynamique moléculaire. Ce point est illustré ci-dessous, à partir de l'étude des propriétés dynamiques de la structure gelée à $t=60$ psec.

Les six premières fréquences propres de cette structure ont respectivement pour valeur

5.8, 6.5, 7.4, 8.5, 9.4, 11.0 cm^{-1} . Celles qui dominent nettement le spectre du mouvement relatif des hélices αB et αC , lors de la trajectoire harmonique calculée à partir des fréquences et des axes propres de cette structure, ont pour valeurs 7.3, 9.3, et 11.0 cm^{-1} . Les deux modes de vibration dont les fréquences sont les plus basses ne contribuent donc pas significativement au mouvement des hélices αB et αC . L'évolution temporelle de l'angle que font les axes de ces deux hélices est montré sur la figure III-18; la comparaison avec les figures précédentes correspondantes montre clairement que l'origine du mouvement ample et lent découvert par Åqvist ne peut être comprise à partir d'études de structures gelées du CTF.

La théorie des modes normaux ne permet donc pas d'expliquer le comportement dynamique relatif des hélices αB et αC du CTF. Qualitativement, ce mouvement est avant tout quasi-périodique, c'est-à-dire à la fois périodique et quelque peu irrégulier, voire chaotique. Or, les résultats issus de l'étude des propriétés dynamiques de structures à basse température, pour lesquelles la théorie des modes normaux donne des résultats exacts, ne mettent pas en évidence ce caractère erratique de la fréquence dominant le mouvement relatif des hélices B et C du CTF; les valeurs calculées sont remarquablement similaires pour les différentes structures étudiées et, de plus, trop élevées, de 8 et 10 cm^{-1} . Ce résultat montre, au passage, que le calcul exact des solutions des équations du mouvement est tout à fait nécessaire pour étudier les propriétés dynamiques d'une macromolécule, qui sont très différentes à température ordinaire de ce qu'elles sont à basse température.

b) Calcul à partir d'états "excités".

Pour étudier les propriétés dynamiques du CTF à température ordinaire, on a calculé les modes de vibration de structures dont l'énergie potentielle n'a pas été préalablement minimisée. On a vu dans la partie II que la projection de la matrice des constantes de force modifie les résultats obtenus lorsque l'énergie potentielle de la structure étudiée n'a pas une valeur minimale. Pour préciser ce point, le calcul des modes de vibration de la

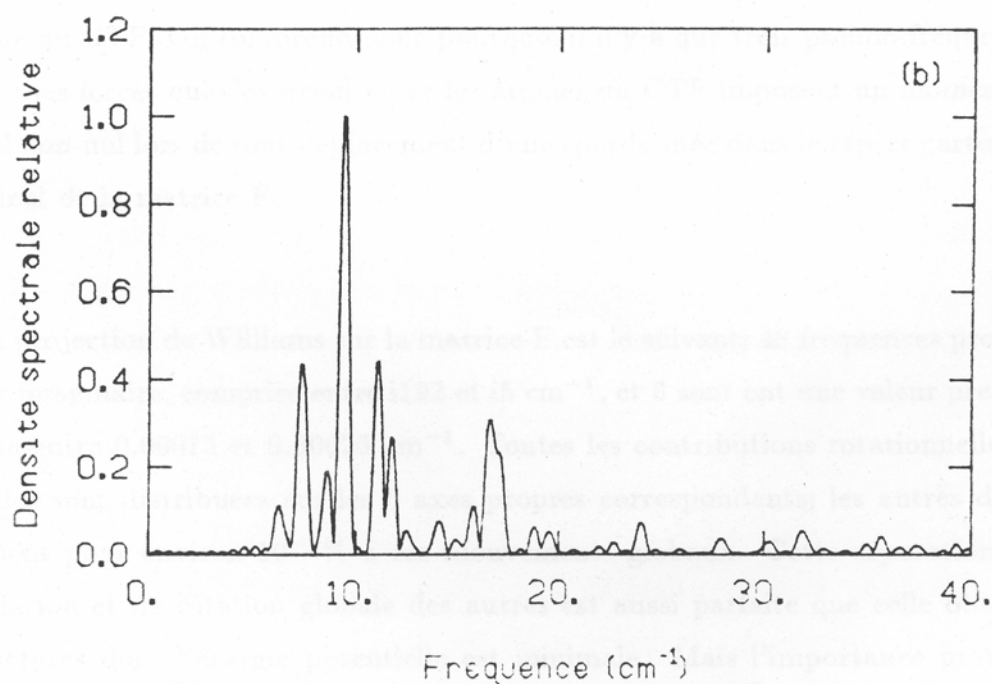
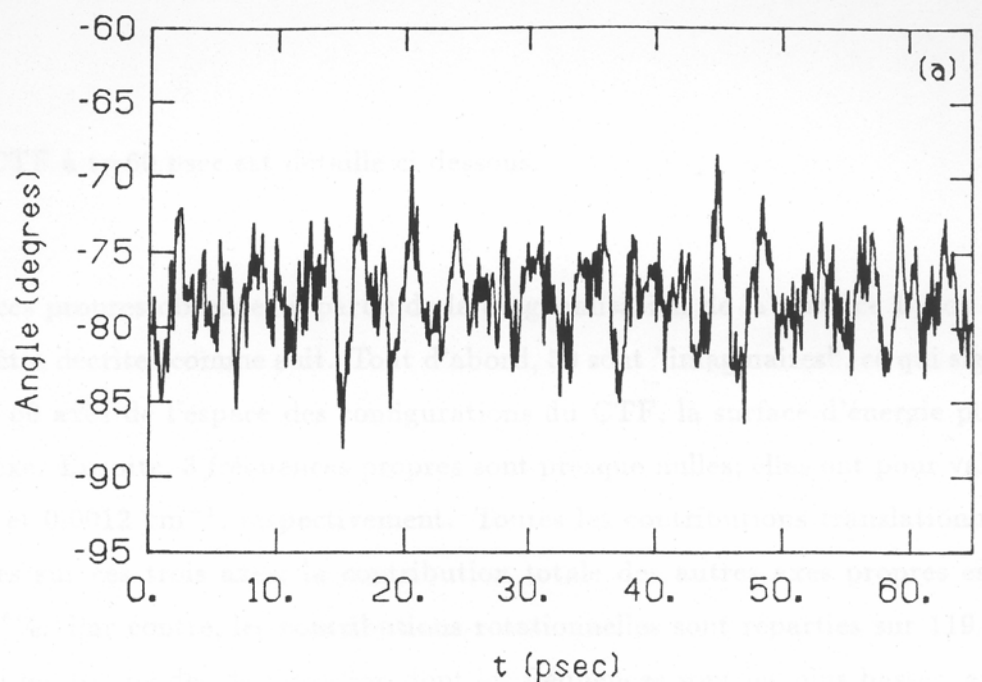


Figure III-18. Trajectoire harmonique calculée à partir des modes de vibration de la structure obtenue après avoir minimisé l'énergie potentielle du CTF lorsque $t=60$ psec. Mouvement relatif des hélices αB et αC : a) évolution de la valeur de l'angle que font les axes de ces deux hélices au cours de la trajectoire du CTF; c) Spectre de puissance relative correspondant.

structure du CTF à $t=60$ psec est détaillé ci dessous. trouvé précédemment.

Les fréquences propres obtenues à partir de la diagonalisation de la matrice \mathbf{F} non projetée peuvent être décrites comme suit. Tout d'abord, 50 sont "imaginaires", ce qui signifie que le long de 50 axes de l'espace des configurations du CTF, la surface d'énergie potentielle est convexe. Ensuite, 3 fréquences propres sont presque nulles; elles ont pour valeurs $i0.002$, 0.0002 et 0.0012 cm^{-1} , respectivement. Toutes les contributions translationnelles sont regroupées sur ces trois axes; la contribution totale des autres axes propres est de l'ordre de $10^{-6}\%$. Par contre, les contributions rotationnelles sont réparties sur 119 axes propres, parmi les 163 modes de vibration dont les fréquences sont les plus basses, et notamment parmi ceux dont les fréquences sont imaginaires, qui contribuent pour 88% à la rotation globale du CTF. On comprend donc pourquoi il n'y a que trois pseudo-fréquences propres nulles : les forces qui s'exercent entre les atomes du CTF imposent un moment de rotation global non-nul lors de tout déplacement d'une coordonnée dans le repère cartésien, pendant le calcul de la matrice \mathbf{F} .

L'effet de la projection de Williams sur la matrice \mathbf{F} est le suivant; 48 fréquences propres ont une valeur imaginaire, comprise entre $i192$ et $i5 \text{ cm}^{-1}$, et 6 sont ont une valeur presque nulle, comprise entre 0.00015 et 0.00026 cm^{-1} . Toutes les contributions rotationnelles et translationnelles sont distribuées sur les 6 axes propres correspondants; les autres directions contribuent pour environ $10^{-7}\%$ à ces mouvements globaux. Cette séparation des axes de translation et de rotation globale des autres est aussi parfaite que celle obtenue pour des structures dont l'énergie potentielle est minimale. Mais l'importance pratique de la projection de Williams n'apparaît pas que pour cet aspect là de la recherche des axes propres d'une structure. Par exemple, les quatre premières fréquences propres positives ont maintenant pour valeurs 2.9 , 3.4 , 6.4 , et 7.3 cm^{-1} , respectivement. Avant la projection, elles avaient pour valeur : 4.6 , 5.0 , 6.5 et 7.3 cm^{-1} , respectivement. D'autre part, l'opération de projection a fait pivoter les axes propres correspondants d'environ 30° . D'autres axes ont été plus affectés encore, notamment ceux le long desquels la surface d'énergie potentielle est convexe. L'un d'eux, par exemple, a pivoté de près de 70° ; ce qui

signifie qu'il n'a presque plus aucun rapport avec celui trouvé précédemment.

Les modifications introduites par la projection de I.H. Williams sont donc très importantes, notamment en ce qui concerne le calcul des axes des modes de vibration de basse fréquence, lorsque l'énergie potentielle de la structure étudiée n'est pas minimale. Après cette opération, les axes et les fréquences propres obtenues ont un sens physique. On peut vérifier que le long des nouveaux axes propres, la courbure de la surface d'énergie potentielle est bien donnée par la valeur des nouvelles fréquences propres. Ces axes et ces fréquences permettent donc d'exploiter l'information locale, c'est-à-dire la courbure de la fonction énergie potentielle, V , au voisinage de la structure étudiée.

Le long de certains de ces axes, la courbure de la surface d'énergie potentielle est convexe, ce qui n'a aucun équivalent lors de l'analyse classique des modes de vibration d'une structure. Comme par ailleurs on veut étudier le mouvement relatif des hélices αB et αC , et que l'on sait que ce mouvement a des composantes de basse fréquence, on ne prend pas en compte ces axes lors du calcul des trajectoires harmoniques. On écarte aussi, lors de ce calcul, les axes des modes de vibration de haute fréquence (supérieurs à 200 cm^{-1}), qui contribuent très peu aux déplacements atomiques.

Lors du calcul des trajectoires harmoniques, on s'intéresse aux fluctuations des quantités suivies; notons que si le comportement dynamique du CTF était parfaitement harmonique, ces fluctuations seraient de même amplitude que celles obtenues à partir de l'étude d'une structure gelée.

Dans le cas de la structure instantanée à $t=60\text{ psec}$ (figure III-19), deux fréquences dominent le spectre du mouvement relatif des hélices αB et αC ; elles ont pour valeurs respectives $3,2$ et $6,4\text{ cm}^{-1}$. Ce résultat signifie qu'il existe des axes de l'espace des configurations le long desquels des déplacements collectifs d'atomes entraînent un mouvement relatif de ces deux hélices. Il n'était pas évident, a priori, que des mouvements collectifs puissent

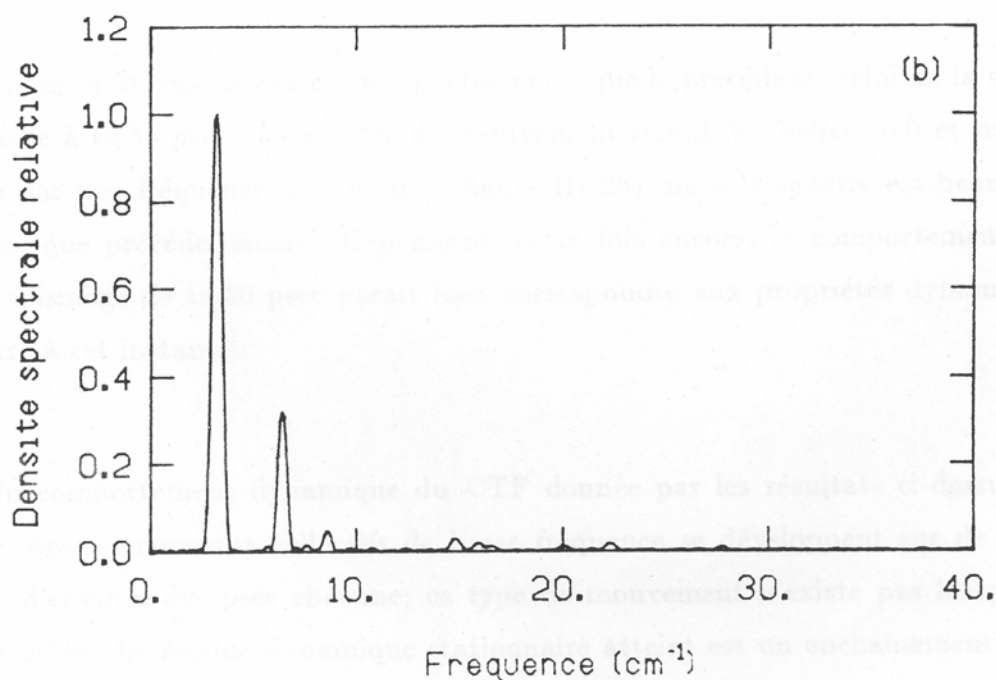
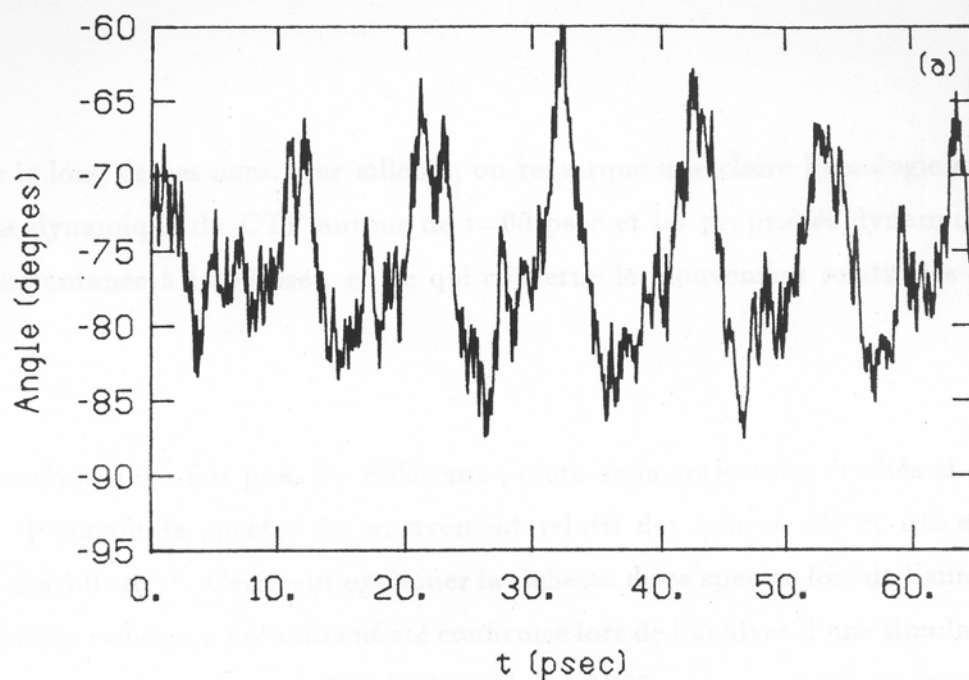


Figure III-19. Trajectoire harmonique calculée à partir des modes de vibration de la structure du CTF à $t=60$ psec. Mouvement relatif des hélices αB et αC : a) évolution de la valeur de l'angle que font les axes de ces deux hélices au cours de la trajectoire du CTF; c) Spectre de puissance relative correspondant.

se développer le long de ces axes. Par ailleurs, on remarque une claire homologie entre le comportement dynamique du CTF autour de $t=60$ psec et les propriétés dynamiques de la structure instantanée à $t=60$ psec, en ce qui concerne le mouvement relatif des hélices αB et αC .

Ce type d'analyse a été fait pour les différents points de la trajectoire étudiés ci-dessus. La fréquence dominant le spectre du mouvement relatif des hélices αB et αC est très variable, de 1.2 à 8.0 cm^{-1} . Ceci peut expliquer la richesse de ce spectre lors de l'analyse de la simulation; cette richesse a notamment été confirmée lors de l'analyse d'une simulation de 1 nsec , réalisée avec le programme GROMOS (O.Tapia, 1990; communication personnelle).

Nous détaillons ci-dessus un cas moins spectaculaire que le précédant, celui de la structure instantanée à $t=50$ psec. Le spectre du mouvement relatif des hélices αB et αC est alors dominé par une fréquence de 3.9 cm^{-1} (figure III-20), mais le spectre est beaucoup plus complexe que précédemment. Cependant, cette fois encore, le comportement dynamique au voisinage de $t=50$ psec paraît bien correspondre aux propriétés dynamiques de la structure à cet instant.

L'image du comportement dynamique du CTF donnée par les résultats ci-dessus est la suivante : des mouvements collectifs de basse fréquence se développent sur de courtes périodes, d'environ dix psec chacune; ce type de mouvement n'existe pas lorsque la structure est gelée. Le régime dynamique stationnaire atteint est un enchainement complexe de telles périodes. La notion de fréquence instantanée émerge ici naturellement; les méthodes qui permettent d'étudier des variations de fréquence devraient donc être appliquées avec succès à l'étude de la dynamique du CTF, comme, par exemple, celles basées sur l'utilisation des "ondelettes" (L.Daubechier & col., 1986; A.Grossman & J.Morlet, 1987; F.Argoul & col., 1989).

Dans la perspective initiale de notre travail, nous retiendrons des résultats ci-dessus

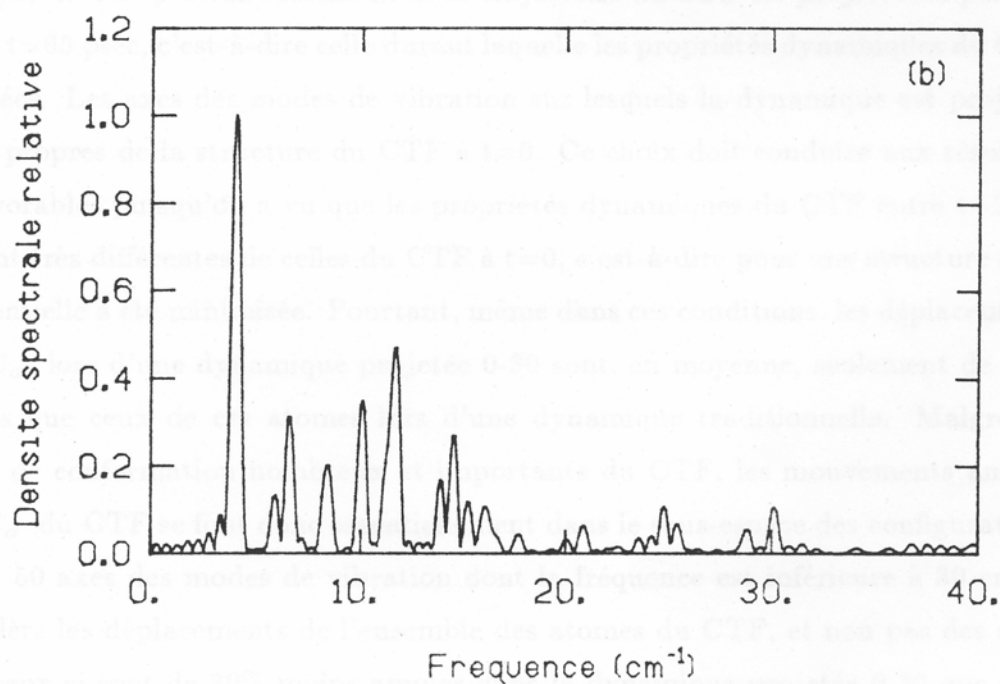
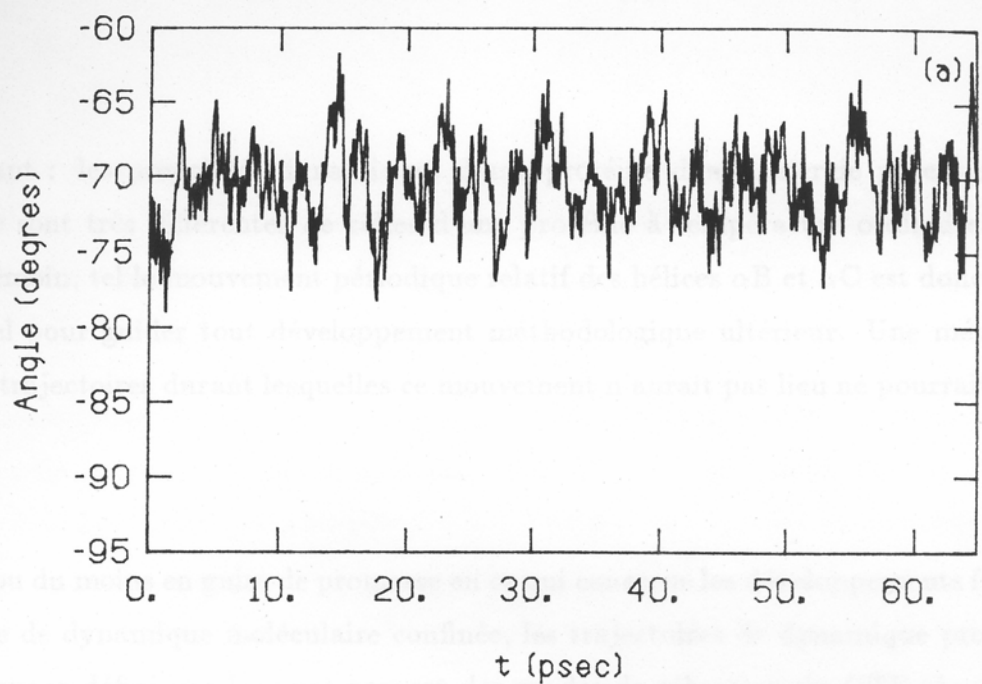


Figure III-20. Trajectoire harmonique calculée à partir des modes de vibration de la structure du CTF à $t=50$ psec. Mouvement relatif des hélices αB et αC : a) évolution de la valeur de l'angle que font les axes de ces deux hélices au cours de la trajectoire du CTF; c) Spectre de puissance relative correspondant.

le point suivant : les propriétés dynamiques d'une protéine dont l'énergie potentielle a été minimisée sont très différentes de celles d'une protéine à température ordinaire. Un phénomène témoin, tel le mouvement périodique relatif des hélices αB et αC est donc tout à fait essentiel pour guider tout développement méthodologique ultérieur. Une méthode calculant des trajectoires durant lesquelles ce mouvement n'aurait pas lieu ne pourrait être validée.

Pour finir, ou du moins en guise de promesse en ce qui concerne les développements futurs de la méthode de dynamique moléculaire confinée, les trajectoires de dynamique projetée sur un sous-espace défini par les axes propres des modes de vibration du CTF (équation 42) sont analysées. Une portion seulement de la trajectoire du CTF est projetée, la portion entre $t=20$ et $t=65$ psec, c'est-à-dire celle durant laquelle les propriétés dynamiques du CTF ont été étudiées. Les axes des modes de vibration sur lesquels la dynamique est projetée sont les axes propres de la structure du CTF à $t=0$. Ce choix doit conduire aux résultats les plus défavorables puisqu'on a vu que les propriétés dynamiques du CTF entre $t=20$ et $t=65$ psec sont très différentes de celles du CTF à $t=0$, c'est-à-dire pour une structure dont l'énergie potentielle a été minimisée. Pourtant, même dans ces conditions, les déplacements des atomes C_α lors d'une dynamique projetée 0-30 sont, en moyenne, seulement de 13% moins amples que ceux de ces atomes lors d'une dynamique traditionnelle. Malgré les changements de conformation nombreux et importants du CTF, les mouvements amples des atomes C_α du CTF se font donc essentiellement dans le sous-espace des configurations défini par les 50 axes des modes de vibration dont la fréquence est inférieure à 30 cm^{-1} . Si l'on considère les déplacements de l'ensemble des atomes du CTF, et non pas des seuls atomes C_α , ceux-ci sont de 30% moins amples dans la dynamique projetée 0-30 que dans la dynamique traditionnelle. Lors d'une dynamique projetée 0-250, ce pourcentage tombe à 6%. En d'autres termes, l'essentiel des déplacements amples des atomes du CTF lors d'une dynamique traditionnelle se fait dans un sous-espace des configurations défini par les 500 axes des modes de vibration dont la fréquence est inférieure à 250 cm^{-1} .

Ce résultat, peu évident à priori, démontre qu'une méthode développée dans l'esprit de

notre méthode de dynamique moléculaire confinée pourrait obtenir des résultats comparables à ceux d'une méthode de dynamique moléculaire traditionnelle. Il importe donc de développer plus avant notre méthode.

CONCLUSION

Une voie prometteuse, dont nous n'avons pas encore fait état, est la fusion de notre méthode avec celle actuellement développée par J.Durup (J.Durup, 1990a; 1990b). Cette méthode repose sur le constat que la structure d'une protéine est organisée en une hiérarchie de sous-unités : atomes, groupe d'atomes, résidus, structures secondaires, etc... Dans la méthode de J.Durup, un système de coordonnées incluant les divers degrés de cette hiérarchie naturelle est utilisé. Ce système est défini de manière arborescente, à partir de vecteurs reliant les centres de gravité de deux vecteurs immédiatement antécédents dans la hiérarchie. Le principe du confinement des déplacements des atomes dans la méthode de J.Durup est de ne contraindre que les niveaux les plus bas de la hiérarchie des coordonnées d'une protéine. Une telle méthode pourrait être combinée avec la nôtre en utilisant des axes propres définis en fonction d'un tel système de coordonnées hiérarchiques. En ce qui concerne le CTF, parmi les coordonnées situées au sommet de la hiérarchie, il y aurait les axes des hélices α et des feuillets β .

En toute rigueur, l'étude théorique des propriétés dynamiques d'une protéine devrait être menée en calculant l'évolution temporelle de sa fonction d'onde. En pratique, une étude aussi rigoureuse n'est pas possible. Par contre, en utilisant les nombreuses données expérimentales disponibles, on peut construire une représentation simplifiée des protéines, c'est-à-dire un modèle qui synthétise l'essentiel de ce que l'on sait sur les protéines. Dans ces modèles, une protéine est en général décrite comme un ensemble d'atomes ponctuels entre lesquels s'exercent des forces dont l'intensité et la direction ne dépendent que de la configuration des atomes de la protéine. A priori, en étudiant ces modèles, on exploite les données introduites lors de leur construction. Cependant, comme ces données ont des origines expérimentales diverses, des informations nouvelles peuvent être tirées de leur confrontation. De plus, suffisamment développés, ces modèles devraient devenir de puissants outils d'aide à l'interprétation des résultats expérimentaux rendant compte des comportements d'ensemble des atomes d'une protéine, tels ceux actuellement décrits en termes de "sous-états conformationnels" (H.Frauenfelder & col., 1988). Cependant, pour permettre de comprendre, par exemple, ce qui différencie deux sous-états conformationnels, il faudrait pouvoir calculer des trajectoires de protéine 10 à 100 fois plus longues que les plus longues des trajectoires calculées à ce jour. Dans le présent travail, nous avons entrepris le développement d'une méthode permettant de calculer de telles trajectoires.

Pour ce faire, nous sommes partis, D.Perahia et moi-même, du constat suivant : les mouvements qui déplacent le plus les atomes d'une protéine sont environ 100 fois plus lents que les mouvements de vibration des liaisons covalentes des protéines. Or, avec une méthode qui permet de maintenir constantes les longueurs des liaisons covalentes d'une protéine, telle celle basée sur l'algorithme SHAKE (J.P.Ryckaert & col., 1977), des trajectoires environ 2 fois plus longues que d'ordinaire sont calculées. Donc, avec une méthode ne calculant que les mouvements qui déplacent le plus les atomes d'une protéine, des trajectoires environ 100 fois plus longues que d'ordinaire pourraient être calculées.

La méthode que nous avons développée -la méthode de dynamique moléculaire confinée- permet effectivement de calculer des trajectoires beaucoup plus longues que d'ordinaire.

Par exemple, nous avons intégré les équations du mouvement des atomes de la décaalanine à 300°K avec un pas temporel environ 30 fois plus grand que celui utilisé lors du calcul d'une trajectoire par une méthode de dynamique moléculaire traditionnelle. Comme notre méthode utilise la description que donne la théorie des modes normaux des mouvements amples et lents d'une molécule, et comme les approximations sur lesquelles repose cette théorie sont respectées à très basse température, nous avons testé notre méthode en vérifiant qu'une trajectoire de la décaalanine à très basse température (5°K), obtenue avec notre méthode, diffère très peu d'une trajectoire obtenue avec une méthode traditionnelle, dans des conditions identiques. Cependant, contrairement aux précédentes, les trajectoires de la décaalanine à température ordinaire, obtenues avec la méthode de dynamique moléculaire confinée, sont atypiques. En effet, lors de ces simulations, les atomes se déplacent très peu.

Pourtant, une autre méthode que nous avons développée -la méthode de dynamique moléculaire projetée, parente naturelle de la méthode de dynamique moléculaire confinée- nous a permis de confirmer l'une de nos principales hypothèses, à savoir que les mouvements amples et lents des atomes d'une molécule à température ordinaire sont convenablement décrits, au moins en moyenne, par la théorie des modes normaux.

L'origine de la contrainte qui réduit l'amplitude des mouvements atomiques lorsqu'ils sont calculés par notre méthode est donc fondamentalement anharmonique, c'est-à-dire liée à ce qui différencie qualitativement le comportement dynamique d'une molécule, tel qu'il est prédit par la théorie des modes normaux, du comportement dynamique d'une molécule à température ordinaire. De fait, la méthode de dynamique moléculaire confinée permet de quantifier l'importance des couplages qui existent à température ordinaire entre les mouvements des atomes le long des différents axes de l'espace des configurations définis dans le cadre de la théorie des modes normaux.

Ces contraintes et ces couplages interdisent d'utiliser la méthode de dynamique

moléculaire confinée telle que nous l'avons développée à ce jour pour calculer de longues trajectoires de protéine. Nous cherchons actuellement à modifier notre méthode de façon à éliminer les effets de ces contraintes au fur et à mesure que les atomes se déplacent. Auparavant, nous avons souhaité nous assurer que notre méthode pourrait bien être appliquée au cas des protéines. Pour cela, nous avons étudié un mouvement ample et lent facile à caractériser, le mouvement relatif des hélices αB et αC du fragment C-terminal de la protéine ribosomale L7/L12 d'E.Coli (CTF).

Ce mouvement relatif a été découvert par Åqvist lors de l'analyse d'une trajectoire calculée dans des conditions de température constante, à partir de la fonction énergie potentielle utilisée par le programme GROMOS (W.F.Van Gunsteren & col., 1983; J.Åqvist & col., 1985). Nous avons, O.Tapia et moi-même, retrouvé ce mouvement lors de l'analyse de trajectoires calculées dans des conditions microcanoniques, c'est-à-dire à énergie totale constante, à partir de la fonction énergie potentielle utilisée par les programmes CHARMM (B.Brooks & col., 1983) et X-PLOR (A.Brunger & col., 1989). La comparaison de ces trajectoires, réalisées avec des options méthodologiques très différentes, a permis de montrer que le comportement dynamique d'une protéine est peu sensible aux détails du modèle utilisé pour l'étudier et notamment au choix des options prises pour calculer le terme électrostatique de la fonction énergie potentielle des protéines.

Pour étudier le mouvement relatif des hélices αB et αC du CTF, nous l'avons comparé au mouvement relatif de ces deux hélices tel qu'il est prédit par la théorie des modes normaux, c'est-à-dire à partir des propriétés dynamiques de cette protéine à très basse température. Là encore, nous avons mis en évidence un effet anharmonique. Tandis que la fréquence moyenne de ce mouvement est, à très basse température, comprise entre 8 et 10 cm^{-1} , à température ordinaire cette fréquence varie entre 3 et 8 cm^{-1} , selon les portions de trajectoires analysées. Ce résultat est pour nous l'indication que la prochaine version de notre méthode de dynamique moléculaire confinée devra utiliser les propriétés dynamiques des protéines telles qu'elles peuvent être déduites de l'étude d'une structure à un instant quelconque d'une trajectoire calculée à température ordinaire et non pas à partir de celles

qui peuvent être déduites de l'étude d'une structure gelée.

Lorsque nous serons en mesure de calculer des trajectoires de protéine de plus de 100 nsec, nous pourrons, dans un premier temps, tester de manière fine l'ensemble de la démarche empirique qui sous-tend l'application de méthodes de simulation au cas des protéines. Actuellement, les données expérimentales les plus précises concernent le mouvement de l'atome de fer de la myoglobine lors de périodes d'une centaine de nsec, environ. Il s'agirait donc de retrouver l'évolution de l'amplitude de ce mouvement en fonction de la température, telle qu'elle est observée par spectroscopie Mossbauer.

Ensuite, comme on montre expérimentalement qu'en 100 nsec une molécule de Myoglobine isolée passe par un nombre important de sous-états conformationnels, les méthodes d'analyse des trajectoires de dynamique moléculaire permettront d'isoler et de caractériser chacun de ces sous-états. Il s'agirait là d'un apport important à la compréhension de la dynamique des protéines, et donc d'un pas important vers la compréhension des mécanismes impliqués dans la fonction des protéines.

REFERENCES

- Anderson, C.M., Stenkamp, R.E., McDonald, R.C., & Steitz, T.A. (1978) *J. Mol. Biol.* **123**, 207.
- Anderson, C.M., Zucker, F.H., & Steitz, T.A. (1979) *Science* **204**, 375.
- Aqvist, J., Van Gunsteren, W.F., Leijonmarck, M., & Tapia, O. (1985) *J. Mol. Biol.* **183**, 461.
- Aqvist, J. (1987) *Thèse, Université d'Uppsala (Suède)*.
- Argoul, F., Arnéodo, A., Grasseau, G., Cagne, Y., Hopfinger, E.J., & Frisch, U. (1989) *Nature* **338**, 51.
- Austin, R.H., Beeson, K.W., Eisenstein, L., Frauenfelder, H., & Gunsalus, I.C. (1975) *Biochemistry* **14**, 5355.
- Bauminger, E.R., Cohen, S.G., Nowik, I., Ofer, S., & Yariv, J. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 736.
- Bennett, W.S., & Steitz, T.A. (1980) *J. Mol. Biol.* **140**, 211.
- Berendsen, H.J.C., & Van Gunsteren, W.F. (1984) "Molecular liquids dynamics and interactions", Barnes, A.J., & col. eds, Reidel Publ. Co. Dordrecht, N.L, NATO ASI Series **C135**, 475.
- Berendsen, H.J.C., Postma, J.P.M., Di Nola, A., Van Gunsteren, W.F., & Haak, J.R. (1984) *J. Chem. Phys.* **84**, 3684.
- Brooks, B., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., & Karplus, M. (1983) *J. Comp. Chem.* **4**(2), 187.

- Brooks, B., & Karplus, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6571.
- Brown, K.G., Erforth, S.C., Small, E.W., Peticolas, W.L. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1467.
- Brunger, A., Karplus, M., & Petsko, G.A. (1989) *Acta Cryst.* **A45**, 50.
- Califano, S. (1976) "Vibrational states", Wiley.
- Case, D.A., & Karplus, M. (1979) *J. Mol. Biol.* **132**, 343.
- Chothia, C., & Lesk, A.M. (1984) *J. Mol. Biol.* **174**, 175.
- Colonna-Césari, F., & Tapia, O. (1990) *en préparation*.
- Constanciel, R., & Tapia, O. (1978) *Theoret. Chem. Acta* **48**, 75.
- Crawford Jr., B.L., & Fletcher, W.H. (1951) *J. Chem. Phys.* **19**(1), 141.
- Cusack, S., Smith, J., Finney, J., Tidor, B., & Karplus, M. (1988) *J. Mol. Biol.* **202**, 903.
- Cusack, S. (1990) *en préparation*.
- Daubechier, L., Grossman, A., & Meyer, Y. (1986) *J. Math. Phys.* **27**, 1271.
- Futrell, R.P., & McGuffey, D.J. (1971) *J. Chem. Phys.* **52**, 285.
- Dennison, D.M. (1940) *Rev. Mod. Phys.* **12**, 175.
- Go, N., Noguti, T., & Nishikawa, T. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3696.
- Dijkstra, B.W., Van Nes, G.J.H., Kalk, K.H., Brandenburg, N.P., Hol, W.G.J., & Drenth, J. (1982) *Acta Crystal. B* **38**, 793.
- Durup, J. (1990a) *J. Chim. Phys.*, sous presse.
- Durup, J. (1990b) *soumis au J. Phys. Chem.*

- Dwivedi, A.M., & Krimm, S. (1982) *Macromolecules* **15**, 186.
- Eklund, H., Samama, J.P., Wallen, L., Braenden, C.I., Akeson, A., Jones, T.A. (1981) *J. Mol. Biol.* **146**, 561.
- Elber, R., & Karplus, M. (1987) *Science* **235**, 318.
- Englander, S.W. (1975) *Ann. N. Y. Acad. Sci.* **244**, 10.
- Englander, S.W., & Englander, J.J. (1978) *Methods in Enzymol.* **49**, 24.
- Fanconi, B., & Peticolas, W.L. (1971) *Biopolymers* **10**, 2223.
- Fanconi, B., Small, E., & Peticolas, W.L. (1971) *Biopolymers* **10**, 1277.
- Fehlhammer, H., Bode, W., & Huber, R. (1977) *J. Mol. Biol.* **111**, 415.
- Fermi E. (1931) *Z. Phys.* **71**, 250.
- Frauenfelder, H., Petsko, G.A., & Tsernoglou, D. (1979) *Nature* **280**, 558.
- Frauenfelder, H., Parak, F., & Young, R.D. (1988) *Ann. Rev. Biophys. Biophys. Chem.* **17**, 451.
- Futrelle, R.P., & McGinty, D.J. (1971) *J. Chem. Phys.* **12**, 285.
- Go, N., Noguti, T., & Nishikawa, T. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3696.
- Go, N. (1989) *Proteins* **5**(2), 104.
- Goldstein, H. (1950) "Classical mechanics", Addison-Wesley, Reading, Massachusetts.
- Goldstein, M. (1969) *J. Chem. Phys.* **51**, 3728.

Grossman, A. & Morlet, J. (1987) "Mathematics and Physics. Lecture on recent results", ed. Streit L., World Scientific, Singapour.

Gudkov, A.T., Tumanova, L.G., Gongadze, G.M., & Bushuev, V.N. (1980) *FEBS letters* **109**, 34.

Gudkov, A.T., & Gongadze, G.M. (1984) *FEBS letters* **176**, 32.

Gurd, F.R.N., & Rothgeb, T.M. (1979) *Adv. Prot. Chem.* **33**, 73.

Gwinn, W.D. (1971) *J. Chem. Phys.* **55**(2), 477.

Hockney, R.W., & Eastwood, J.W. (1981) "Computer simulations using particles" McGraw-Hill, New York.

Holbrook, J.J., Liljas, A., Steindel, S.J., & Rossmann, M.G. (1975) *The Enzymes* **11**, 191.

Hoy, A.R., Mills, I.M., & Strey, G. (1972) *Mol. Phys.* **24**, 1265.

Huber, R., & Bennett, W.S. (1983) *Biopolymers* **22**, 261.

Hvidt, A., & Linderstrom-Lang, K.U. (1954) *Biochim. Biophys. Acta* **14**, 574.

Iben, E.T., Braunstein, D., Doster, W., Frauenfelder, H., Hong, M.K., Johnson, J.B., Luck, S., Ormos, P., Schulte, A., Steinbach, P.J., Xie, A.H., & Young, R.D. (1989) *Phys. Rev. letters* **62**(16), 1916.

Itoh, K., & Schimanouchi, T. (1970) *Biopolymers* **9**, 383

Kabsch, W. (1976) *Acta Crystal. A* **32**, 922.

Karplus, M., & McCammon, J.A. (1981) *CRC Crit. Rev. Biochem.* **9**, 243.

Karplus, M., & McCammon, J.A. (1983) *Ann. Rev. Biochem.* **53**, 263.

- Karplus, M., & McCammon, J.A. (1986) *Pour La Science* **104**, 42.
- Katz, H., Walter, R., & Somorjai, R.L. (1979) *Computers & Chemistry* **3**, 25.
- Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C., & Shore, V.C. (1960) *Nature* **185**, 422.
- Koshland, D.E. (1958) *Proc. Nat. Acad. Sci. USA* **44**, 98.
- Koshland, D.E. (1960) *Adv. Enzymol.* **22**, 45.
- Koteliansky, V.E., Domogatsky, S.P., & Gudkov, A.T. (1978) *Eur. J. Biochem.* **90**, 319.
- Krimm, S. (1983) *Biopolymers* **22**, 217.
- Lackowicz, J.R., & Weber, G. (1973) *Biochemistry* **12**, 4171.
- Leijonmarck, M., Erikson, S., & Liljas, A. (1980) *Nature* **286**, 824.
- Leijonmarck, M., & Liljas, A. (1987) *J. Mol. Biol.* **195**, 555.
- Levitt, M. (1972) *Thèse, Université de Cambridge (Grande Bretagne)*.
- Levitt, M., Sander, C., & Stern, P.S. (1983) *Int. J. Quant. Chem. : Quant. Biol. Symp.* **10**, 181.
- Levitt, M., Sander, C., & Stern, P.S. (1985) *J. Mol. Biol.* **181**, 423.
- Levy, R.M., Perahia, D., & Karplus, M. (1982) *Proc. Nat. Acad. Sci. USA* **79**, 1346.
- Liljas, A., & Leijonmarck, M. (1983) "Supramolecular structure and function", Pifat, G., & Herak, J.N., eds, Plenum Press, New York.
- Linderstrom-Lang, K.U., & Schellman, J.A. (1959) *The Enzymes* **1**, 443

- Loncharich, R.J., & Brooks, B. (1989) *Proteins* **6**(1), 32.
- Luer, C.A., & Wong, K.P. (1980) *Biochemistry* **19**, 176.
- Maasen, J.A., & Moller, W. (1981) *Eur. J. Biochem.* **115**, 279.
- Marquis, D.M., & Fahnestock, S.R. (1980) *J. Mol. Biol.* **142**, 161.
- McCammon, J.A., & Harvey, S.C. (1987) "Dynamics of proteins and nucleic acids", Cambridge University Press.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., & Teller, F. (1953) *J. Chem. Phys.* **21**, 1087.
- Mills, I.M. (1969) *Chem. Phys. letters* **3**, 267.
- Mills, I.M. (1974) *Theor. Chem.* **1**, 111.
- Moller, W., Groene, A., Terhorst, C. & Amons, P. (1972) *Eur. J. Biochem.* **25**, 5.
- Moller, W. (1974) "Ribosomes", p711, Namura, M., Tissieres, A., & Lengyel, P., eds, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Moller, W., Schrier, P.I., Maasen, J.A., Zantema, A., Schop, E., & Reinalda, H. (1983) *J. Mol. Biol.* **163**, 553.
- Noguti, T., & Go, N. (1982) *Nature* **296**, 776.
- Noguti, T., & Go, N. (1983) *J. Phys. Soc. Japan* **52**(9), 3283.
- Noguti, T., & Go, N. (1985) *Biopolymers* **24**, 527.
- Osterberg, R., Sjoberg, B., Liljas, A., & Petterson, I. (1976) *FEBS letters* **66**, 48.

- Papousek, D., & Aliev, M.R. (1982) "Molecular Vibrational-Rotational Spectra", Elsevier.
- Parak, F., & Knapp, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7088.
- Pariseau, M.A., Suzuki, I., & Overand, J. (1965) *J. Chem. Phys.* **42**, 2335.
- Pauling, L. (1960) "The Nature of the Chemical Bond", Cornell University Press, New York.
- Perahia, D., Levy, R.M., & Karplus, M. (1990) *Biopolymers* **29**, 645.
- Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G., & North, A.C.T. (1960) *Nature* **185**, 416.
- Perutz, M.F., & Matthews, F.S. (1966) *J. Mol. Biol.* **21**, 199.
- Peticolas, W.L. (1979) *Methods in Enzymol.* **61**, 425.
- Petterson, I., & Kurland, C.G. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 4007.
- Philipps, D.C. (1966) *Scientific American* **215**, 78.
- Powell, M.J.D. (1977) *Math. Program.* **12**, 241.
- Pulay, P. (1969) *Mol. Phys.* **17**, 197.
- Pulay, P. (1977) *Modern Theor. Chem.* **4**, 153.
- Rabolt, J.F., Moore, W.H., & Krimm, S. (1977) *Macromolecules* **10**, 1065.
- Remington, S.J., Wiegand, G., & Huber, R. (1982) *J. Mol. Biol.* **158**, 111.
- Ryckaert, J.P., Ciccotti, G., & Berendsen, H.J.C. (1977) *J. Comp. Phys.* **23**, 327.

- Sanejouand, Y.H. (1985) *Mémoire de D.E.A., Université d'Orsay (France)*.
- Sanejouand, Y.H., Perahia, D., & Karplus, M. (1990) *en préparation*.
- Sanejouand, Y.H., & Tapia, O. (1990) *en préparation*.
- Simonson, T. (1990) *Thèse, Université d'Orsay (France)*.
- Sloane, C.S., & Hase, W.L. (1977) *J. Chem. Phys.* **66**(4), 1523.
- Smith, J., Kuczera, K., Tidor, B., Doster, W., Cusack, S., & Karplus, M. (1989) *Physica B* **156-157**, 437.
- Streett, W.B., Tildesley, D.J., & Saville, G. (1978) *Mol. Phys.* **35**(3), 639.
- Swaminathan, S., Ichiye, T., Van Gunsteren, W.F., & Karplus, M. (1982) *Biochemistry* **21**, 5230.
- Tainer, J.A., Getzoff, E.D., Beem, K.M., Richardson, J.S., & Richardson, J.S. (1982) *J. Mol. Biol.* **160**, 181.
- Tapia, O. (1980) "Quantum theory of chemical reactivity", vol. 2, p25, Riedel, Dordrecht, Holland.
- Tapia, O. (1990) *à paraître*.
- Tyson, J., Claassen, H.H., & Kim, H. (1971) *J. Chem. Phys.* **54**, 3142.
- Tobias, D.J., & Brooks III, C.L. (1988) *J. Chem. Phys.* **89**(8), 5115.
- Van Agrhoven, A.J., Maasen, J.A., Schrier, P.I., & Moller, W. (1975) *Biochem. Biophys. Res. Commun.* **64**, 1184.
- Van Gunsteren, W.F., & Karplus, M. (1982) *Macromolecules* **15**(6), 1528.

- Van Gunsteren, W.F., Berendsen, H.J.C., Hermans, J., Hol, W.G., & Postma, J.P.M. (1983) *Proc. Nat. Acad. Sci., USA* **80**, 4315.
- Verlet, L. (1967) *Phys. Rev.* **159**, 98.
- Warshel, A., & Russel, S.T. (1984) *Quarterly Rev. Biophys.* **17**(3), 283.
- Whitlow, M., & Teeter, M.N. (1986) *J. Amer. Chem. Soc.* **108**(23), 7163.
- Wilkinson, J.H., & Reinsch, C. (1971) "*Handbook for Automatic Computation*", Springer-Verlag, Berlin.
- Williams, I.H. (1983) *J. Mol. Struct.* **94**, 275.
- Williams, R.J.P (1979) *Biol. Rev. Cambr. Philos. Soc.* **54**, 389.
- Wilson, E.B. (1939) *J. Chem. Phys.* **7**, 1047.
- Wilson, E.B. (1941) *J. Chem. Phys.* **9**, 76.
- Wilson, E.B., Decius, J.C. & Cross, P.C. (1955) "*Molecular Vibrations*", McGraw-Hill, New York.

RESUME

La plasticité des protéines joue un rôle majeur dans l'expression de leur fonction. Or, les déplacements amples de groupes d'atomes à l'intérieur des protéines sont souvent difficiles à étudier expérimentalement. Par exemple, on ne sait dire ce qui distingue entre eux les sous-états conformationnels mis en évidence par Frauenfelder. Pour préciser l'interprétation de ce type de donnée expérimentale, les méthodes de dynamique moléculaire seraient idéales si le calcul de trajectoires de protéines d'environ 100 nsec était possible. La méthode de dynamique moléculaire confinée que nous avons développée repose sur la description que donne la théorie des modes normaux des mouvements amples et lents d'une protéine. Elle permet de calculer des trajectoires beaucoup plus longues que d'ordinaire. Cependant, un comportement anharmonique méconnu perturbe le déroulement des trajectoires calculées par notre méthode, même dans le cas d'un polypeptide, une décaalanine, ne subissant aucun changement de conformation. Pour préciser les voies de développement ultérieur de notre méthode, nous avons consacré la troisième partie de notre thèse à l'étude d'un mouvement ample et lent à l'intérieur d'une petite protéine, le fragment C-terminal de la protéine ribosomale L7/L12.

MOTS CLEFS

Théorie des modes normaux; Projection de Williams; Anharmonicité; Fréquences et directions propres; Axes des modes de vibration; Dynamique moléculaire confinée; Décaalanine; Protéine ribosomale L7/L12.